



École de gestion

Sujet :

**Étude comparative entre les méthodes du pari ordinaire  
et les choix expérimentaux discrets pour le calcul  
d'algorithmes de QALY**

Par

Pape SénySy

Mémoire présenté en vue de l'obtention du grade de maîtrise en économie

Octobre 2017

© Pape SénySy, 2017

Ont été évaluées par un jury composé des personnes suivantes :

Jie He..... Directeur de recherche

Thomas Poder..... Codirecteur de recherche

Jason Robert Guertin..... Autre membre du jury

Mémoire accepté le .....

## Table des matières

Introduction .....	4
I. <i>Littérature</i> .....	6
1. <i>LE PARI ORDINAIRE OU STANDARD GAMBLE (SG)</i> .....	8
A. <i>La théorie de l'utilité espérée</i> .....	8
B. <i>Avantages du SG dans la prévision du QALY</i> .....	9
C. <i>Les inconvénients du pari ordinaire</i> .....	11
2. <i>Le classement contingent et le TTO</i> .....	15
3. <i>Le choix expérimental discret (DCE)</i> .....	16
A. <i>La théorie de l'utilité aléatoire</i> .....	16
B. <i>Les avantages de la méthode par rapport au SG</i> .....	17
C. <i>Les inconvénients</i> .....	18
II. <i>Méthodologie</i> .....	21
1. <i>La sélection des états de santé</i> .....	22
2. <i>L'enquête</i> .....	24
3. <i>Les données</i> .....	30
4. <i>Le modèle</i> .....	39
III. <i>Résultat</i> .....	44
1. <i>SG</i> .....	44
2. <i>DCE</i> .....	52
IV. <i>Discussion</i> .....	58
1- <i>L'instrument</i> .....	58
2- <i>La méthode d'évaluation</i> .....	58
<b>3- Le modèle</b> .....	60
Conclusion.....	67

# Introduction

Le but des services et programmes sanitaires est d'améliorer la santé des individus. Lorsque la santé est bonne, elle améliore le bien-être et c'est la raison pour laquelle au cours de ces dernières décennies une importance accrue a été donnée au développement des méthodes de mesures quantitatives de l'utilité des états de santé. L'objectif de ce développement est d'identifier et de quantifier l'ampleur des changements liés à la santé au sein de la société suite à l'application d'un traitement. Cependant, les ressources dont on dispose en vue du financement des différents programmes de santé ne sont disponibles qu'en quantité limitée. Si l'on désire être efficace, cela impose un choix qui sous-entend une comparaison de ces interventions. On peut les comparer au niveau du coût qui a été mobilisé pour les financer, mais surtout au niveau des résultats qu'ils ont permis d'atteindre. Et dans ce cas, la littérature fait appel à la combinaison de deux dimensions qui sont *l'amélioration de la qualité de vie* et *la durée* qui sont regroupées dans un instrument appelé le QALY (*Quality-Adjusted Life Year*) ou *année de vie pondérée de la qualité*. Ce dernier facilite la mesure de l'utilité de ces programmes sanitaires en se basant sur des méthodes comme le pari ordinaire, le *marchandage-temps* (*Time Trade-Off*, TTO) ou compromis sur la durée, les *échelles graduées VAS* (*Visual Analogic Scale*, VAS), le *choix expérimental discret* ou encore la méthode du classement (*Ranking*). Ces méthodes ont pour objectif la prise en compte des préférences des individus dans la détermination d'un algorithme de QALY.

La méthode du pari ordinaire ou "*Standard Gamble*" (SG) est utilisée en économie pour accorder une valeur aux préférences des individus dans une situation risquée. Par exemple, elle sert en économie de la santé à donner une valeur à l'utilité accordée à un état de santé sur une échelle normée. Cette méthode présente cependant certaines limites, car elle est associée à la théorie de l'utilité espérée qui est contestée quant à sa capacité à permettre d'effectuer des choix dans une situation avec risque (Van Houtven et al. 2007). Le SG utilise la probabilité qui n'est pas évaluée de façon linéaire telle que décrite par la théorie (Bleichrodt 2002) contrairement à la méthode des choix expérimentaux discrets standard (« *Discrete Choice Experiment* (DCE) ») qui fait ici abstraction de toutes ces notions. Cependant, cette dernière ne respecte pas souvent l'hypothèse des alternatives non

cohérentes (Craig et al. 2007). Rappelons ici que le bien-être ou encore le résultat d'un traitement sont des concepts très relatifs pour les individus. Ainsi une méthode de détermination du QALY diverge d'une autre dans sa manière d'inclure les préférences des individus dans ce calcul. C'est la raison pour laquelle le QALY d'un état de santé est souvent différent d'une méthode à une autre dans les études empiriques.

Ainsi le but de cette étude est de comparer la méthode SG, et celle du DCE dans le calcul d'un QALY. Aucune comparaison directe entre ces deux méthodes n'a été effectuée dans les études citées, ni d'ailleurs à notre connaissance dans d'autres études visant à établir un mode de calcul de QALY. Dans notre recherche, nous proposons de comparer l'efficacité du SG à celle du DCE en utilisant à cet effet les cartes de choix utilisées par le SF6Dv2 (annexe 1). Ce dernier est un outil qui permet de définir un état de santé à partir de 6 dimensions distinctes (fonctionnement physique, limitation des activités, fonctionnement social, douleur, santé mentale et vitalité). (Voir section 2 pour plus de détails). Le SF6Dv2 est tiré du SF36v2 (Cécile et al. 2007) un outil internationalement validé dans le domaine de la mesure de la qualité de vie et qui est utilisé pour calculer des QALY. Il est en cours de validation dans plusieurs pays dont le Royaume-Uni et le Canada. Ainsi, notre étude permet de comparer ces deux méthodes afin de déterminer laquelle permet de fournir le meilleur système de pondération aux 6 dimensions considérées dans le questionnaire SF6Dv2. Ce faisant, elle permettra d'établir quelle méthode reflète au mieux les préférences des Québécois pour différents états de santé et ainsi tenir compte de leur spécificité dans la construction de QALY.

À la première section sur la littérature concernée, nous avons recensé un ensemble d'articles qui utilisent ces deux méthodes dans le calcul de QALY. Selon nos recherches, il n'existe pour l'instant que des études théoriques qui effectuent la comparaison des méthodes SG et DCE, qui mettent en relief les avantages et les limites des deux méthodes (Van Houtven et al. 2007; Starmer 2000). Notre objectif consistera à réaliser une première base de comparaison empirique entre ces deux méthodes.

La deuxième section est consacrée à l'explication des méthodologies. Nous évoquons ici le choix d'un échantillon représentatif des états de santé, l'enquête, quelques statistiques

descriptives de l'échantillon des répondants et enfin les modèles et tests de comparaisons utilisés. Les différents états de santé sont issus des 18 750 possibilités offertes par les choix de réponses au questionnaire SF6Dv2. Dans la mesure où il n'est pas possible d'évaluer la totalité des 18 750 états de santé possibles, 120 états sont retenus lors d'une procédure de sélection orthogonale. Cette procédure permet l'identification d'un modèle pour estimer la valeur de tous les 18 750 états de santé (« orthogonal main effects design »). Les états choisis ont fait l'objet d'une étude en ligne. Durant cette enquête, les répondants sont aléatoirement répartis en deux groupes, correspondant chacun à l'une des méthodes. Les données recueillies de l'enquête sont par la suite appliquées à différents modèles que l'on a présentés à la section 2 sur la méthodologie. L'efficacité de chacune de ces méthodes sera par la suite mesurée par un ensemble de tests de comparaison, dont le pouvoir explicatif du modèle ( $R^2$ ) le critère d'information d'Akaike, l'erreur absolue moyenne (MAE), le test de Jarque et Béra et celui de Ljung-Box.

La dernière section présente les résultats de l'étude. De toutes les spécifications utilisées pour le SG notamment dans le modèle effet fixe, on note une bonne aptitude de prévision des états de santé malgré une absence de significativité qui est enregistrée dans les coefficients de la fonction de régression. Pour la méthode du DCE, pratiquement l'ensemble des coefficients estimés sont significatifs (24/25 coefficients pour le modèle LOGIT à coefficients aléatoires). Cependant cette significativité est accompagnée de problèmes soulevés par le critère de MAE qui rend la prévision des états de santé difficile pour cette méthode. À ce stade, la méthode du SG qui dispose d'un avantage dans la prévision semble être la meilleure alternative. Quelques ajustements sont par la suite faits dans le DCE avant la conclusion définitive sur les deux méthodes. De ce fait, la méthode choisie sera recommandée dans la prévision du QALY des états de santé. Son utilisation aboutira à une analyse coût-utilité plus adéquate et permettra de financer un programme de soins sanitaires qui répondra au mieux aux préférences du consommateur pour un meilleur niveau de bien-être.

## *I. Littérature*

Le concept de QALY a été approché en premier dans les travaux de (Klarman et al. 1968). Mais le terme de « Quality-Adjusted Life Years » a été employé en premier dans les travaux de Zeckhauser et al. (1976). Le QALY est un instrument économique qui est utilisé en santé pour faciliter le choix d'une thérapie parmi plusieurs. Il est obtenu en faisant le produit de la qualité de vie par la durée. Une année vécue dans la bonne santé à la suite d'une intervention est évaluée à un QALY de 1, si l'intervention ne réussit pas et conduit à la mort le QALY sera égal à zéro par contre si elle permet de rallonger la durée de vie ou d'éviter la mort au cours d'une année, au prix d'un handicap, il est compris entre 0 et 1. Il s'inscrit dans le cadre de maximisation de l'utilité (CES 2003). L'une des motivations qui encouragent l'utilisation du QALY est qu'il permet de rationaliser et d'optimiser le peu de ressources dont dispose le domaine médical privé ou public compte tenu de la forte demande. La mise en place d'une thérapie fait appel à une étude au préalable de coût efficacité. On pouvait être beaucoup plus tenté par une approche coût-bénéfice comme il est de coutume en économie. La différence majeure entre l'analyse coût-bénéfice et le coût-efficacité réside dans l'unité de mesure des résultats. Avec l'approche coût-utilité, les résultats sont exprimés en ratios constitués par un coût rapporté à un jugement utilitaire de la qualité de vie qui est en général issu d'une préférence. Cependant, les individus ont des préférences en matière de santé qui sont plutôt diversifiées. Ainsi, pour respecter ce principe de maximisation de l'utilité, le QALY doit prendre en compte les préférences que les individus ont à l'égard des états de santé. Une préférence exprime le choix d'un individu qui peut être basé sur un critère objectif ou subjectif. Le calcul du QALY se base ainsi sur des méthodes qui tentent de capter ces préférences individuelles. Il en existe plusieurs qui sont généralement réparties en deux groupes :

Celles qui incluent les préférences individuelles à travers une *utilité cardinale*. Ces méthodes se concentrent beaucoup plus à la valeur approximative que les individus accordent aux différents états de santé. On peut citer le SG, la méthode du TTO, etc. Le SG présente deux alternatives aux individus; un état de santé et un traitement. Il leur demande par la suite de choisir l'un d'entre eux compte tenu des différentes chances de succès du traitement jusqu'à la situation d'indifférence. Le TTO quant à lui présente deux états de santé aux individus puis fait varier la durée de vie jusqu'à ce que l'individu soit indifférent entre les deux. Ces méthodes sont bien détaillées dans la section méthodologie.

Celle de *l'utilité ordinale* s'attarde quant à elle sur la façon dont les individus vont classer les différents états de santé. C'est le cas de la méthode DCE. Cette dernière demande à l'individu de choisir entre deux états de santé sur la base de ses préférences.

Avec le QALY on peut appréhender la préférence des individus en matière de santé et les thérapies permettant d'y accéder de manière plus efficace. Cependant le choix entre les différentes méthodes permettant de capter au mieux les préférences des individus dans ce calcul est questionnable. Notre objectif à travers cette littérature sera donc de décrire ce qui a été fait dans le calcul du QALY, avec les méthodes SG et DCE.

### 1. LE PARI ORDINAIRE OU STANDARD GAMBLE (SG)

Il est considéré comme la méthode standard dans la mesure de l'utilité des états de santé pour des raisons normatives, mais aussi du fait que le risque est un aspect important dans un contexte de décision médicale (Bleichrodt et al. 1999). Il reste jusqu'à nos jours la seule méthode qui se base dans une situation risquée. En effet, elle inclut dans sa méthodologie une notion probabiliste présente dans les différents résultats d'un traitement et les incorpore dans le calcul comme stipulé par la théorie de l'utilité espérée.

#### *A. La théorie de l'utilité espérée*

L'inclusion des préférences individuelles dans le calcul du QALY est une étape importante. C'est la raison qui explique l'existence de ces diverses méthodes. Elles ont trouvé comme base axiomatique de calcul d'optimisation la théorie de l'utilité espérée de Von Neumann Morgenstern (1944) dont le modèle vise à décrire le comportement individuel en univers incertain (Feeny et Torrance 1989). Cette théorie, appliquée à notre étude sur la santé, signifie que les résultats d'une intervention sont associés à une valeur d'utilité. Chaque résultat est multiplié par sa fréquence d'occurrence et une fois sommée, elle fournit une utilité espérée que chaque individu rationnel tentera de maximiser. Cette méthode correspond assez bien à la description que la théorie donne au choix individuel en univers incertain. Elle détermine l'utilité d'un état de santé en essayant d'approximer le maximum de risque qu'un individu est prêt à prendre pour atteindre un niveau de santé parfait. Elle incorpore le risque, mais cela suppose que les individus soient neutres vis-à-vis de ce risque. Si cette hypothèse est valide, la méthode devrait parvenir à de bonnes performances dans la détermination du QALY d'un état de santé. Dans le cas contraire, les résultats



devraient être considérés avec précaution. De plus, selon certains auteurs comme Van Houtven, l'utilité espérée ne reflète pas les préférences des individus dans une situation impliquant le risque (Van Houtven et al. 2007). En d'autres termes, chacune des méthodes que nous verrons dans cette argumentation dispose de ces avantages et inconvénients.

### *B. Avantages du SG dans la prévision du QALY*

Le SG a été instauré dans le processus de détermination du QALY pour capter au mieux le comportement des individus. Si tel est le cas, on devrait enregistrer une plus grande précision dans le calcul du QALY. Dans ces études que l'on aura à citer, le QALY ne fait pas toujours l'objet de calcul (en tout cas pas de façon directe), mais un contrôle de la performance de la méthode peut être fait au niveau des paramètres de la fonction de régression. La variable dépendante de cette fonction est souvent le QALY qu'un individu accorde à un état. L'article de Brazier et al. (2002) est une référence pour les études portant sur le QALY. Les auteurs ont fait une application du SF6D pour la première fois au Royaume-Uni. Les résultats qu'ils ont enregistrés diffèrent selon le type de modèle choisi. Par exemple une décroissance plus accentuée des paramètres du modèle MCO que celui à effet fixe. Le modèle à effet aléatoire utilisé après la réalisation du test de Hausman renferme 17 coefficients significatifs pour un pouvoir explicatif de 0,204 et un MAE de 7%. Avec ces résultats, ils ont conclu que l'ensemble des modèles ont tendance à surestimer la valeur des états faibles et à sous-estimer celle des états à valeur forte. Une approche identique a été faite à partir de la population japonaise sur la base du SG à l'aide de modèles à effet aléatoire et MCO (Brazier et al. 2009). Les résultats aboutissent à des conclusions un peu différentes de ceux trouvés dans l'étude appliquée au Royaume-Uni (Brazier et al. 2002). En effet dans l'étude appliquée au Japon, les modèles utilisés (qui sont les mêmes) ont des performances bien en deçà de ceux appliqués au Royaume-Uni que ce soient dans la significativité des variables, erreurs de prévisions ou biais. À noter que cette étude de 2002 a tenté d'autres spécifications, mais en vain. Cela aboutira au contraire à des prédictions biaisées (t tests) associées à une erreur de mesure plus considérable. La difficulté à prédire le QALY ne découle pas forcément du SG. Il peut être issu de l'utilisation du SF6D qui venait de faire son apparition dans ce domaine. En effet, les données provenant de l'enquête faite sur la population du Royaume-Uni renferment une certaine bimodalité et une certaine asymétrie faisant que les mesures de tendance centrale

restent difficilement applicables. Ils ont aussi trouvé qu'elles renferment une certaine hétérogénéité avec une différence qui est enregistrée entre les répondants, mais aussi à l'intérieur de chaque répondant.

Une autre application du SF6D a été faite aux États-Unis par Craig et al. (2013). La principale variante demeure l'utilisation des paires permutées qui restent une composante de la méthode du DCE dans la mesure où elle demande aux répondants de faire un choix par rapport à des alternatives qui varient suivant deux attributs dans le but de réduire la difficulté. L'utilisation d'un probit conduit en somme à des résultats assez similaires que ceux trouvés dans Brazier et al. (2002) dans la significativité des paramètres même si la valeur des coefficients n'est pas tout à fait la même. Un coefficient de corrélation de 0,941 a été trouvé entre les QALY des deux études. L'utilisation du modèle aléatoire épisodique fut critiquée (Salomon et Menzies 2011). Ils s'opposent à l'emploi de ce modèle qui alloue une durée similaire de dix ans qui va à l'encontre des principes du TTO. Ces deux études bien qu'étant un peu différentes dans la conception aboutissent à des résultats similaires. Même si elles rencontrent des problèmes qui leur sont un peu spécifiques, elles parviennent assez bien à des résultats satisfaisants. Pour chaque type de problème, les auteurs ont tenté tant bien que mal d'apporter des corrections. Par exemple dans l'étude de Brazier et al. (2002) Les auteurs ont assemblé (en grappe) les données par individu et estimé des modèles à l'échelle individuelle et agrégée pour faire face à l'hétérogénéité. Une même asymétrie a été enregistrée dans l'étude menée aux USA, mais pour ce cas, elle est imputable, selon les auteurs, à l'effet Hawthorne (lorsque les répondants modifient leur comportement en réponse à l'étude) et certains biais de désirabilité sociaux (tendance à répondre de manière favorable<sup>1</sup>). En dépit de ces corrections, on se trouve pour le moment dans l'incapacité d'expliquer certaines limites comme la sure/sous-estimation de certaines valeurs du QALY et certaines inconsistances dans les modèles faisant que certaines questions méritent d'être soulevées. Une première peut être reliée à la théorie, notamment l'hypothèse très forte sur laquelle repose la méthode, c'est-à-dire la neutralité des individus vis-à-vis du risque. À en juger par certains résultats ci-dessus, on se demande si elle est respectée. L'autre

---

<sup>1</sup>Cela doit être très limité, car il n'y a pas de réponses attendues, comme dans une étude de VAP où l'on s'attend à ce que les gens disent oui ou non. Ici, on parle d'états de santé.

interrogation est relative à la procédure du SG. Par ailleurs, ce dernier est la seule méthode qui présente de façon claire l'existence d'un traitement dans sa procédure d'évaluation.

### *C. Les inconvénients du pari ordinaire*

#### *a. La théorie*

Certaines irrégularités notamment liées aux coefficients des paramètres pourraient s'expliquer par quelques-unes des limites associées à la théorie. Bleichrodt (2002) souligne 3 raisons qui font que les individus violent l'hypothèse de l'utilité espérée. Il s'agit pour le SG de la pondération des probabilités, de l'aversion pour la perte et de la courbure de la fonction de durée.

- Si les individus violent la théorie de l'utilité, c'est en partie liée au fait qu'ils n'évaluent pas la probabilité de façon linéaire comme prédit par la théorie. La fonction de pondération devrait plutôt s'écrire de la manière suivante :

$W(p) \cdot U(q_1, t) + (1 - W(p)) \cdot U(q_2, t)$ . C'est-à-dire une combinaison non linéaire de l'utilité contrairement à la théorie. Les fonctions d'utilité représentent chaque résultat du traitement dans le SG. Ce qui fait qu'au point d'indifférence la valeur qui sera associée à l'état de santé sera la fonction  $W(p)$  plutôt que  $(p)$ .  $W(p)$  représente ici une fonction de probabilité définie sur un ensemble réel positif. Ainsi si  $W(p)$  est concave ( $W$  croissant et  $w''$  négatif), le QALY est biaisé vers le bas. S'il est convexe ( $W$  croissant et  $W''$  positif), il est biaisé vers le haut. Dans l'étude de Bleichrodt et al. (1999), les auteurs ont fait une étude qui a comme but de savoir si l'utilisation de l'utilité non espérée peut améliorer la validité descriptive du SG sans réduire ou sacrifier son statut normatif. Ils ont utilisé un modèle de rangement des alternatives dépendantes de l'utilité. Il s'agit pour le QALY de contrôler la cohérence lorsque cette fonction de pondération est considérée. Ils ont trouvé que le QALY pris avec la fonction de pondération probabiliste est significativement plus cohérent que celui qui n'en tient pas compte. Mais que ce résultat suppose que la courbure de la fonction d'utilité ne soit pas prise en considération. Si tel n'est pas le cas (prise en compte de cette courbure), les améliorations en matière de cohérence sont moindres. Suivant les études de Camerer (1994), Kahneman (1992) et Gonzalez (1996), l'estimation du paramètre de cette fonction de probabilité

montre que plusieurs répondants se situent sur la partie convexe de cette fonction. En d'autres termes, l'application de la théorie de l'utilité espérée crée un biais dirigé vers le haut.

- Une autre raison expliquant la déviation des individus vis-à-vis de l'utilité espérée est l'aversion pour la perte. En effet au point d'indifférence, les résultats des états de santé sont évalués en considération de pertes et de gains (Kahneman, 1979). Van Houtven et al. (2007) vont plus loin dans cette considération qu'a le SG par rapport au risque. En effet, le risque est un aspect central dans la plupart des contextes de décision médicale et c'est la raison pour laquelle le SG a été avancé comme le procédé standard de mesure de l'utilité de la santé (Bleichrodt 1999). S'il dispose de l'avantage de la représentation des décisions dans le cadre incertain, plusieurs écrits critiquent l'utilisation de l'utilité espérée pour l'unique raison qu'il ne représente pas le comportement observé des individus en univers incertain (STARMER 2000). On aura alors besoin d'avoir une idée sur l'impact de l'utilité non espérée sur le maximum de risque toléré par un individu. En effet, la décision de choix d'une intervention requiert davantage de connaissances sur le compromis entre le bénéfice que rapporte ce dernier et le maximum de risque que le concerné est capable d'accepter. Si le SG n'en tient pas compte, cela signifie que ses résultats peuvent être biaisés. L'étude de Van Houtven et al. (2011) Appréhende le QALY dans le cadre d'une représentation de l'utilité non espérée et voit l'impact de cette représentation sur le risque maximum toléré. Ils sont aidés dans cette tâche par l'utilisation de trois modèles d'approche qui ont tous une différente appréhension du problème. Le premier considère le risque pris comme une variable catégorique plus précisément « adverse event risk level ». Il aboutit à une défaillance dans 2 parmi les 3 types de risques de la théorie de l'utilité espérée. La conclusion est que le modèle considéré doit tenir compte de l'hétérogénéité des individus vis-à-vis du maximum de risque qu'ils sont en mesure de tolérer et qui varie suivant le type de risque. Le second modèle est la considération d'une simple fonction de pondération. Et la dernière est basée sur le cadre d'un modèle de rangement dépendant de l'utilité. Cette dernière fournit des résultats plus robustes. Leur conclusion est que le contrôle des trois facteurs suivants détermine la fonction de probabilité qui doit

être accordée à chaque alternative dans l'utilité espérée : le type de risque, le maximum de risque toléré par l'individu et le compromis risque/bénéfice. À croire que l'hypothèse de la neutralité vis-à-vis du risque n'est pas vérifiée.

b. La neutralité vis-à-vis du risque

La valeur d'un même état de santé peut être différente selon la méthode de détermination des préférences utilisée. L'approche de Wakker et al. (1995) consiste à essayer d'expliquer la différence d'utilité entre le SG et le TTO. Ils ont ajusté la méthode du TTO à travers l'attitude de l'individu par rapport au risque (risk attitude) pour voir si cela pouvait éliminer la différence d'utilité des deux méthodes. Les résultats ont montré une utilité du SG significativement plus importante que celle du TTO non ajusté. Comme la majorité des répondants ont fait preuve d'une aversion à l'égard du risque, le TTO ajusté renferme des utilités plus fortes que le TTO non ajusté et se rapproche davantage du pari ordinaire. L'aversion au risque explique dans cette étude une des raisons pour laquelle un même état de santé peut avoir diverses valeurs suivant les méthodes d'approche. Cela pourrait expliquer aussi certaines divergences des résultats que l'on a vus précédemment. C'est la raison pour laquelle il est important d'avoir des modèles plus flexibles capables de déceler l'attitude envers le risque de la personne à qui le choix s'applique. À noter que cette étude est faite sur un échantillon de 30 patients atteints d'un cancer pour un total de 4 états de santé. Les résultats doivent donc être pris avec précaution. Mais cela n'empêche pas ce problème de neutralité par rapport au risque de persister.

Mise à part la théorie, la différence d'utilité d'une méthode par rapport à l'autre peut aussi se situer au niveau de la procédure d'évaluation.

c. La valorisation des états

Les états ne sont pas évalués tous de la même façon dans la méthode du SG. Si le but du SG est d'incorporer les préférences des individus en matière de soins de santé, il doit être en mesure d'évaluer l'ensemble des états de la même manière. On se trouve dans l'incapacité de savoir l'impact que cette différence d'évaluation aura dans les prévisions de QALY. Rowen et al. (2014), ont mis sur pied une approche qui permet d'évaluer tous les états de la même manière. Ils présentent aux répondants la possibilité de choisir cette fois entre deux traitements : l'un où les états et leur chance de succès sont connus, et un autre où seulement les résultats sont connus. Il sera question pour eux de choisir la probabilité

d'échec du second traitement qui leur rendraient indifférent entre les deux. Ainsi tous les états seront évalués de la même manière et contenus sur l'échelle (0-1). Dans l'étude de Craig et al. (2013), seulement 5 % des états sont évalués d'une manière différente par rapport aux autres états de santé qui ont fait l'objet d'étude. Ce chiffre tourne autour de 7% dans Brazier et al. (2002). Mais le problème réside sur le fait qu'il n'y a pas une seule façon de traiter ce problème. Il en existe plusieurs qui n'aboutissent pas tout le temps à la même conclusion.

Cet aspect ne permet pas d'expliquer toutes les difficultés de la méthode liées aux inconsistances ou à la différence dans les résultats des études citées puisque certaines insuffisances sont toujours liées à la méthode choisie ou à l'étude même (la résultante de la manière d'échantillonnage, si oui ou non il est représentatif de la population ou tout simplement d'une différence culturelle). Et toutes ces limites que nous venons de citer font du SG une méthode dont la mise en place demande un travail assez fastidieux associé à une difficulté à la fois pour le chercheur et son échantillon.

#### d. La difficulté dans le SG

Un reproche que l'on peut faire à cette méthode est bien évidemment sa difficulté d'exécution. Les études empiriques ont montré que sur un échantillon représentatif, seulement une infime partie est en mesure de comprendre cette méthode. Même si les concernés sont favorables à l'utilisation, l'expérience a montré qu'ils ne le comprenaient pas pour autant et les résultats sont à prendre avec beaucoup de précautions (Ratcliffe et al. 2011). La difficulté peut être diverse, résultante de plusieurs facteurs et enregistrée à deux niveaux.

##### ➤ Au niveau de la composition l'échantillon

Il n'est pas sûr que le point de vue des adultes et des adolescents soit la même en matière de QALY. De même que les états les mieux évalués par les jeunes adolescents ne sont pas les mêmes que ceux des adultes (Saigal et al. 2003). Ratcliffe et al. (2011), vont plus loin que l'étude de Saigal. En effet, elle et ses homologues montrent que l'échantillon composé d'adultes et d'adolescent de moins de 13 ans comprend mieux la méthode DCE best worstscaling (qui est une composante de la méthode DCE) avec une aptitude à justifier les réponses, alors que seulement 12 % des répondants de la méthode du SG semblent comprendre

cette dernière. Ceci est dû à la difficulté à traiter avec la probabilité, la considération de toutes les dimensions d'un état ou une combinaison des deux (Ratcliffe et al. 2011). Le concept de certains états de santé « mort subite » rend aussi la tâche difficile pour les jeunes.

➤ **Au niveau de la procédure de la méthode**

L'une des méthodes permettant de trouver la situation d'indifférence avec le SG est celle du pingpong. Elle a été développée par une équipe de l'Université McMaster en Ontario (Furlong et al 1990; Torrance et Feeny 1989). Durant cette procédure, on fait varier la probabilité de réussite dans le traitement jusqu'à ce que l'individu soit indifférent entre les alternatives, en commençant avec 100 % santé parfaite et 0 % mort, puis 10 % santé parfaite et 90 % mort, etc. (voir annexe 1). La méthode du pingpong constitue à elle seule un travail fastidieux pour celui qui le met en place et par la suite demande un effort cognitif important pour l'individu. Le manque de compréhension peut expliquer la divergence des résultats de QALY entre le SG et le DCE.

Ainsi la mise en application de cette technique exige davantage d'efforts pour être appliquée par rapport à d'autres méthodes telles que la méthode des choix expérimentaux et lorsque toutes ces limites sont agrégées, elles sont en mesure d'expliquer les inconsistances dans les estimations ou la difficulté de prévision. Nous verrons dans les parties qui suivent les alternatives qui furent proposées.

## 2. Le classement contingent et le TTO

Les techniques de collecte de données ordinales comme le Ranking (méthode de classement des états de santé) ou le VAS (Visual Analogic Scale) renferment un certain nombre d'avantages par rapport aux méthodes comme le SG ou le TTO. Même si ces méthodes ne disposent d'aucun fondement théorique, ils font preuve d'une facilité de compréhension et d'administration sans compter d'une plus grande fiabilité correspondante à la diminution de l'erreur de mesure. Le TTO où la méthode du compromis sur le temps fut en premier utilisée en évaluation économique des états de santé dans le but de réduire la difficulté que présentait le SG. Cette méthode consiste à présenter deux profils d'état de

santé, dont la santé parfaite. Le but est de faire varier la durée de survie dans l'état parfait jusqu'à ce que l'individu soit indifférent entre les deux situations. La facilité du Ranking semble plus marquante dans la tranche de la population associée à un niveau d'alphabétisation limité. Pour cette tranche de la population, la tâche que propose le SG ou le TTO peut être moins plaisante par rapport à celle que les auteurs proposent sur le Ranking. Craig et al. (2009) tentent de mesurer la capacité des méthodes ordinales à traiter avec les données cardinales. Au total, des données provenant de 8 pays (Slovénie, Argentine, Danemark, Japon, Hollande, Espagne Royaume-Uni, États unis) sont retenues et une comparaison sera faite entre les différentes approches que sont le VAS, TTO et le Ranking.

Les résultats aboutissent sur le fait que la méthode du Ranking est plus corrélée au VAS qu'à celle du TTO avec des coefficients de corrélation élevés allant de 0,61 à 0,91. Les divergences enregistrées dans les résultats sont dues à l'ajustement dans les données. Finalement, la méthode du Ranking parviendra à capter 86 % des variations dans les deux autres méthodes à travers les différents pays. Et c'est la raison pour laquelle les auteurs recommandent son utilisation dans les études à venir. Ce qui justifie l'utilisation du DCE puisque le Ranking constitue une variante de ce dernier.

### 3. Le choix expérimental discret (DCE)

La conséquence à laquelle on pouvait s'attendre pour faire suite aux limites du SG est bien évidemment l'utilisation d'une méthode capable de simplifier davantage la prise en compte des préférences dans le QALY. C'est ainsi que la méthode des choix expérimentaux discrets offre une approche alternative dans l'évaluation économique en matière de santé. À l'origine, ce modèle n'était pas développé à des fins d'évaluation économique impliquant des choix en matière de santé, mais plutôt dans le domaine du transport et du marketing. Son aptitude à traiter avec les données, le système de rangement dont il dispose associé à sa flexibilité, lui a valu son utilisation dans la sphère économique en matière de santé (Louvrière et al. 2010). Il se base sur la théorie de l'utilité aléatoire.

#### *A. La théorie de l'utilité aléatoire*

Le DCE considère une classe de modèle de choix du consommateur bâti sur la base de la théorie de l'utilité qui a été développée par Thurstone en 1927 connue sous le nom de l'utilité aléatoire (the Random utility theory). Pour connaître la valeur d'une loterie suivant



cette théorie, on transforme les résultats en utilité puis on les classe conformément à nos préférences. Notons que les modèles de choix basés sur la théorie de l'utilité aléatoire s'inscrivent dans la lignée des modèles probabilistes, ils se basent sur le principe de maximisation de l'utilité parce que les individus choisissent l'alternative qui a la plus grande utilité pour eux. La théorie de l'utilité aléatoire stipule que chaque personne qui est confrontée à un choix entre diverses alternatives dispose d'une utilité qui renferme deux composantes distinctes :

Une composante déterministe; le choix d'un état par exemple pourrait s'expliquer par la présence des niveaux des attributs qui définissent cet état de santé, dans la composition de l'alternative ou des caractéristiques de l'individu.

Une composante aléatoire qui reflète les éléments de subjectivité, les erreurs d'évaluation liées à l'asymétrie de l'information ou à la rationalité limitée du concerné. Les psychologues assument que l'individu a une appréhension de mesure imparfaite donc il est probable que certains des choix ne soient pas rationnels. Le DCE appliqué à cette théorie fournit de bons résultats. Salomon et McCabe (2008) ont démontré que l'utilisation de l'approche ordinale parvenait à de meilleurs résultats dans la prévision du QALY des états de santé. À l'aide de diverses spécifications sur la durée, l'application du DCE peut devenir un puissant outil dans l'évaluation économique des états de santé (Norman et al. 2014). Une conclusion partielle de Robinson et al. (2014) fut la capacité du DCE dans la prévision de valeurs plus exactes, même si ce dernier est associé au risque. Hakim et al. (1999) sont parvenus à conclure avec l'aide d'un modèle multiplicatif que le DCE parvenait à de résultats acceptables en matière de capacités prédictives.

### *B. Les avantages de la méthode par rapport au SG*

Le DCE fut proposé en partie pour venir à bout des limites du TTO et elle parvient à éviter la majorité des désavantages que l'on reproche à la méthode du SG.

#### ➤ La difficulté

Elle est moins importante pour le chercheur et les répondants du fait que la probabilité n'apparaît plus. Aucune difficulté liée à la probabilité ou à l'hypothèse de neutralité du risque n'est observée. Il n'y a aussi plus lieu de processus du Pingpong.

#### ➤ La valorisation des états

Dans la mesure où la méthode demande un simple classement des états basé sur les préférences personnelles, tous les états de santé sont évalués de la même manière.

➤ La théorie

Dans la méthode du SG, on a montré que la théorie de l'utilité espérée ne reflète pas le comportement des individus en univers incertain. Le DCE est une méthode capable de le tester à l'aide de la qualité des prévisions basées sur différents modèles avec diverses spécifications. Robinson et al. (2014) ont comparé des résultats obtenus à partir du DCE appliqués au modèle d'Utilité non Espérée (UNE) et celui de l'Utilité espérée (UE). Ils ont trouvé que l'estimation dans le modèle DCE basé sur le risque utilisé dans le cadre du modèle UNE associe des valeurs d'utilité plus faibles aux états que la traditionnelle UE (modèles UNE et EU sont ici associés au DCE). Entre autres, le contrôle de la fonction de pondération probabiliste (qui n'est disponible que sur le DCE) aboutit à une pondération différente de la forme linéaire qui est la base de l'UE appliquée dans le SG. L'utilisation d'un paramètre de contrôle pour vérifier la nature de cette pondération aboutit à un résultat supérieur à l'unité. Ce qui signifie que la probabilité est surestimée, et explique ainsi les valeurs élevées dans le modèle UE. L'utilité calculée sur la base du modèle UE est donc biaisée vers le haut. D'où la ressemblance entre les modèles  $SG_{UE}$  et  $DCE_{UE}$ . D'autres travaux, Houtven et al. (2011) ont trouvé que l'utilité qui découle de la méthode  $DCE_{UNE}$  est en général plus faible que celle du  $DCE_{UE}$ .

La méthode du DCE semble être une alternative plus simple et flexible aux techniques antérieures.

Mais on verra dans ce qui suit que son utilisation n'est pas exempte de tous reproches.

*C. Les inconvénients*

*L'hypothèse de l'indépendance des alternatives non cohérentes*

Même si l'hypothèse peut avoir des significations diverses selon les domaines d'application, elles ont toutes un point commun qui est de rationaliser le comportement de l'agent dans des situations qui implique des préférences additionnées ; dans notre situation, cela pourrait être vu comme l'aspect multidimensionnel que renferment nos états. L'axiome suivant résume de façon brève cette hypothèse.

Supposons le choix qui porte sur des alternatives IJ dépourvues du syndrome de dominance. Lorsque l'individu choisi I alors que J lui est accessible dans la dimension (IJ), l'apparition d'une nouvelle alternative X qui modifie la dimension précédente en (I J X) ne doit en aucun cas modifier le choix précédent c'est-à-dire I préféré à J. La théorie dit en somme que le choix entre deux options demeure inchangé si l'on introduit une troisième. Le choix préexistant entre IJ ne doit pas être influencé par X qui n'est en aucun cas pertinent dans la relation entre I et J. parmi les 8 pays présents dans l'étude de Craig et al. (2009), 4 ne peuvent rejeter l'hypothèse de l'indépendance des alternatives non cohérentes. Ce que nous devons retenir dans ce test de Hausmann permettant de déceler le respect ou non de cette hypothèse (IIA), c'est qu'il a été réalisé à partir d'un logit conditionnel dont la taille de l'échantillon est diminuée pour permettre l'estimation de paramètres cohérents. Ce qui fait qu'une partie de ces résultats sont attribuables à la variation de la taille de l'échantillon plutôt que l'hypothèse de l'IIA. Mais le même test réalisé par (McCabe et al. 2006) a rejeté cette hypothèse (IIA). Même si la méthode parvient à éviter certaines limites liées au SG et que les résultats sont en faveur de son utilisation dans l'évaluation économique des états de santé, l'absence de cette hypothèse (IIA) est une limite importante. Si cette hypothèse n'est pas respectée, il est probable que l'évaluation des attributs devient dépendant du contexte de choix, ainsi ils ne sont pas évalués de la même manière (Green and Srinivasan 1990).

*a. Une difficulté dans la mise à l'échelle*

Une particularité de la méthode DCE est qu'elle ne permet pas de contenir les états sur une échelle normée nécessaire à l'analyse coût-utilité. Dans le contexte où nous voulons avoir une échelle de comparaison entre les différents programmes d'intervention en santé susceptibles d'être financés, divers procédés sont entrepris pour venir atténuer ce problème.

- i. Utilisation de la mort subite comme un paramètre du modèle à estimer (Brazier et al. 2002)
- ii. Utilisation de la valeur des états jugée comme étant pire que la mort par le TTO (Ratcliffe et al. 2009)
- iii. Mapping qui consiste à utiliser la structure du TTO dans le DCE (Rowen et al. 2014)

- iv. Combiner les données des méthodes DCE et TTO dans un modèle hybride (Rowen et al. 2014)
- v. L'introduction d'un certain risque de décès immédiat comme un choix (Karni2009; Craig et al 2013)
- vi. L'ajout d'un choix d'une durée de survie comme attribut ( $DCE_{TTO}$ ).

Les quatre premières adaptations ont fait l'objet d'une comparaison par Rowen et al. (2014). Dans cette étude il sera question de comparer différentes méthodes qui ont pour mission de contenir les états de santé sur une échelle normée pour faciliter l'analyse coût-utilité. Les méthodes plus anciennes (i, ii) sont comparées aux nouvelles méthodes (iii et iv). Les résultats ont été plus en faveur des nouvelles méthodes qui produisent des estimations plus semblables au TTO et les auteurs supportent l'idée que la combinaison du DCE et du TTO est plus adéquate dans la procédure de mise à l'échelle des états. Ce qui signifie que les nouvelles méthodes concordent mieux entre elles.

Mais la valeur que l'on accorde à un état de santé est inextricablement liée à la durée de vie (Gafni 1994). Raison pour laquelle nous optons pour la dernière approche. Dans les études citées, on assume que le QALY est le produit entre l'utilité de l'état de santé (qualité de vie) et la durée. Sans la capacité d'explorer ce compromis de qualité de vie/durée (qui est perdue lorsque la durée est omise), cette interprétation des données est impossible justifiant ainsi l'utilisation du  $DCE_{TTO}$ .

Le TTO ne sera pas entièrement copié par DCE, en effet ce dernier lui emprunte sa dimension temporelle pour les raisons que nous avons vues dans le paragraphe précédent et s'abstient du compromis de la durée dans la mesure où il ne demande qu'un simple classement des états définis cette fois par les différents attributs ajoutés de la dimension temporelle, d'où le nom de  $DCE_{TTO}$ . Elle a été appliquée dans les études de Nick Bansback (2012) pour faire l'objet de comparaison par rapport au TTO. Les auteurs ont trouvé qu'elle dispose de certains avantages par rapport à la méthode TTO. Le  $DCE_{TTO}$  va favoriser le fait d'évaluer tous les états de santé de la même manière sans passer par une étape de transformation ou d'évaluation particulière. Ce qui aura un impact dans la robustesse des résultats. Les auteurs montrent aussi une similitude des deux méthodes dans l'évaluation des états de santé en dépit de la différence de spécification, logit conditionnel à effet aléatoire pour le  $DCE_{TTO}$  et paramétrique à effet aléatoire pour le TTO. Par contre, une

grande divergence se trouve dans les états considérés comme étant faibles. L'inclusion des participants n'ayant pas trop cerné la tâche du DCE a une influence peu significative selon les auteurs de l'étude. Ce qui témoigne de la capacité du DCE à réduire le biais lié à l'exclusion de certains répondants (Bansback 2012). Il se base sur le modèle de l'utilité aléatoire. L'utilisation du modèle DCE<sub>TO</sub> semble être bénéfique, mais elle cache certaines irrégularités sur un état de santé en particulier (mort subite) (Norman et al. 2014).

*b. La situation à l'état de la mort*

L'inclusion de la durée donne au QALY la possibilité de prendre la valeur nulle qui est considérée comme étant l'état de la mort. Il est ici défini comme étant un état à part entier dont la durée de vie est nulle. Les études empiriques montrent que les individus ont de la difficulté à traiter avec cette situation. Ratcliffe et al. (2014) avec son étude appliquée aux jeunes et adolescents a trouvé que cette partie de la population a beaucoup de mal à traiter avec les situations impliquant la mort subite. Angela Robinson en 2014 a trouvé que le meilleur modèle parmi ceux qui ont fait l'objet d'étude (SG<sub>UE</sub>, DCE<sub>UE</sub>, DCE<sub>UNE</sub>) en matière de prévision est celui qui considère différentes sources de variation du terme d'erreur notamment celle qui se rattache à la mort (Robinson et al. 2014). En effet, la décomposition de la variance du terme d'erreur a révélé une grande fluctuation vis-à-vis de la mort (variance bien plus importante que celle qui est standard ou découlant de la dominance). Dans la mesure où le DCE considère la mort comme un potentiel alternatif, les jeunes et adolescents qui considèrent que toute la vie est digne d'être vécue, quel que soit l'état de santé où l'on se trouve ne choisiront jamais cette alternative. Ce qui constitue une violation de la théorie de l'utilité aléatoire.

Ces deux méthodes (SG et DCE), bien qu'étant différentes dans la conception, disposent de certains avantages et limites qui leur sont spécifiques. Notre apport consistera à les comparer dans la prévision du QALY des différents états de santé. On sera aidé dans cette tâche par l'utilisation du SF6Dv2 qui est appliqué pour la première fois au Québec.

## II. Méthodologie

*1. La forme réduite en 6 dimensions*

Le calcul du QALY nécessite d'utiliser un outil décrivant les états de santé. Cet outil se doit d'être simple et compréhensible pour faciliter son utilisation. Il en existe plusieurs permettant de décrire un état de santé, dont le SF6Dv2 (annexe 2). Ce dernier est dérivé du SF36v2. Il est associé à une moindre difficulté (on réduit le nombre de dimensions à 6 ainsi que les différents niveaux de ces dimensions). Il apporte également des améliorations techniques dans la mesure où il fait la distinction du passage d'un niveau à un autre à travers les différentes dimensions. Ce qui n'était pas le cas du SF36. Le SF6D a été développé par Brazier et al. (1998, 2002, 2004). Chacune des 6 dimensions renferme un total de 4 à 6 déclarations ordonnées (niveaux). *Un état est ainsi composé d'un ensemble de déclarations dont chacune provient d'une dimension distincte.* Une version améliorée du SF6D, soit le SF6Dv2 est en cours de validation dans plusieurs pays dont le Royaume-Uni, et le Canada. Une étude de comparaison du SF6Dv2 a été faite par Poder et al. (2017) au Canada. Elle avait pour but de voir si administrer le SF6Dv2 avec la totalité (10 questions nécessaires pour le SF36) ou une partie du SF36 (soit 6 questions reformulées) pouvait changer les résultats. La conclusion indique une faible différence selon le mode d'administration. Ce qui justifie l'utilisation du SF6Dv2 (tiré de la version courte du SF36) dans cette étude comme outils permettant de définir un état de santé. Il renferme un total de 18750 états de santé. Pour une utilisation effective du questionnaire, il a fallu choisir un échantillon représentatif.

### 1. La sélection des états de santé

Il existe autant d'états qu'il y a de combinaisons entre les différents niveaux des attributs, soit  $5 \times 5 \times 5 \times 6 \times 5 \times 5$  états de santé pour un total de 18750. Ce nombre se verra multiplié par 4 avec l'addition de la dimension temporelle qui comporte 4 niveaux (1, 4, 7, et 10 ans). Ce qui fait que tous les états ne peuvent pas faire objet d'une évaluation. On se doit de choisir un échantillon des états qui est représentatif, équilibré et optimal. Les techniques de conception expérimentale (Louvière, Hensher et Swait 2000) et les logiciels expérimentaux et de conception SAS ont été utilisés pour obtenir une conception orthogonale, composée de 120 états. Ce nombre est le minimum d'état qu'on peut obtenir tout en gardant un critère d'efficience. Certains états ont quand même été volontairement omis de cette sélection. Il s'agit de la santé parfaite (111111) du pire état de santé (PITS)

que peut offrir le SF6D (555655). Ces deux états de santé constituent des bases dans la méthode du SG que l'on décrira dans la suite. Certaines combinaisons entre les différents niveaux des attributs ont aussi été contraintes. Il s'agit de combinaisons qualifiées d'improbables. Un état est qualifié d'improbable si par exemple il décrit d'excellentes caractéristiques dans les activités physiques quotidiennes (dimension "activité physique"), et associe des douleurs extrêmes dans la dimension "douleur". Toutes ces contraintes ont été programmées dans le logiciel économétrique SAS pour que celui-ci en prenne compte au moment du choix des états de santé et assurer l'efficacité des choix.

L'utilisation du logiciel économétrique SAS nous a permis de choisir un échantillon d'état de santé pour notre étude. Le choix des états est fait en considérant tous les niveaux de chaque dimension du SF6Dv2 y compris la durée à l'aide de la commande « %mktruns (5 5 5 6 5 5); ». Cette déclaration des différentes dimensions constitue la première étape du processus de sélection des états. Il permet de prendre en compte l'ensemble des combinaisons possibles. C'est durant cette première étape que l'on a obtenu 120 états de santé comme un échantillon qui est équilibré compte tenu de la taille du SF6Dv2. Cet échantillon est associé à un D-efficacité de 99,85%. Notons qu'à ce niveau seul la taille de l'échantillon est obtenue et non sa composition. Pour savoir les états qui feront partie de ces 120, il a d'abord fallu préciser à notre logiciel économétrique les combinaisons à éviter dans une macro de restriction « %macro res ». C'est ainsi que la combinaison entre le niveau 1 de la dimension *limitation des activités* et les niveaux 5 et 6 de la dimension *douleur* ont été exclues. Il a été de même pour le niveau 1 de la dimension *fonctionnement social* et les niveaux 5 et 6 de la dimension *douleur* ainsi que le niveau 5 des dimensions *limitation des activités sociales*, *fonctionnement* et le niveau 1 de la dimension *douleur*.

Après avoir défini l'ensemble des contraintes liées à la nature des états, on génère 120 combinaisons qui représentent notre échantillon dans cette étude.

On arrive par la suite à l'étape où l'on distingue ces 120 états selon qu'ils soient utilisés dans le SG ou le DCE. Pour le SG, on fixe la durée à 10 ans conformément aux exigences de la méthode. Pour le DCE, on utilise les macros « *%mktlab*, *%mktex* » pour définir la formation en CHOICE SET et une contrainte de non dominance stricte entre les états à l'intérieur de chaque CHOICE SET. Ce dernier est défini comme un pair d'état de santé dans lequel le répondant est appelé à faire un choix. Les 120 états de santé sont maintenant regroupés en 60 CHOICE SET à des fins d'utilisation du DCE. Mais cette contrainte de non-dominance étant assez forte, on s'est retrouvé avec un problème de duplication des états de santé. En effet dans les 60 CHOICE SET, on se retrouvait avec une dizaine d'états de santé dont certains apparaissaient en double ou parfois même en triple. En d'autres termes, tous les 120 états n'étaient pas présents à l'intérieur des CHOICE SET. Ce qui a été à l'origine de la dernière commande « *%mktdups* » qui consistait à faire de sorte qu'un état de santé qui figure parmi les 120 choisis ne soit représenté qu'une et une seule fois dans le CHOICE SET. La commande « *proc print* » est utilisée pour faire apparaître l'ensemble des états choisis à la fin du processus.

Au total 120 états sont retenus pour faire l'objet d'évaluation. Ces 120 états de santé sont présentés en 24 groupes de 5 pour les individus qui évaluent le SG. Pour le DCE, on utilise ces mêmes 120 états, mais cette fois en paires. Chaque répondant au DCE a 10 paires à évaluer. La seule différence entre les états utilisés entre le SG et le DCE réside dans la durée qui est fixée à 10 ans sur le SG et qui varie en fonction des états avec le DCE.

## 2. L'enquête

Les études basées sur le QALY se tournent de plus en plus vers des enquêtes en ligne (Craig et al. 2013) et ce sera le cas pour celui-ci. Après une brève description sur le but de l'étude et ce qu'elle nous permettra de réaliser dans un futur proche (c.-à-d. avoir la meilleure méthode de détermination du QALY pour une application à différents traitements), on commence l'enquête par un ensemble de questions qui visent dans un premier temps à décrire les caractéristiques sociodémographiques et le niveau de satisfaction que les répondants portent à leur état de santé. Par la suite, on leur présente une succession de questions pour essayer de les familiariser à la tâche des deux méthodes. Par exemple, on leur demande de classer 6 états de santé. Ce groupe est composé de l'état de santé parfait,



du pire état de santé, de la mort subite et de trois autres états intermédiaires qui varient entre les répondants. Cette étape les familiarise à la notion de choix par rapport à la description que fait le SF6Dv2 pour les différents états de santé. On évalue s'ils sont en mesure de comprendre la notion de probabilité en leur montrant des images qui représentent un certain taux de succès d'un traitement. À partir de cet instant, l'échantillon est réparti aléatoirement en deux groupes : une partie pour les questions relatives au SG et une autre partie pour les questions du DCE.

Pour le SG : les 120 états sont répartis en 24 groupes de 5 et distribués de façon aléatoire aux individus. Chaque individu évalue un groupe de 5 états de santé.

La méthode du pari ordinaire propose aux individus le choix entre les alternatives suivantes :

- Alternative 1 : vivre dans un état de santé imparfait (intermédiaire) pendant 10ans et certain de se réaliser.
- Alternative 2 : la possibilité de bénéficier d'un traitement risqué qui aboutit soit à l'état de santé parfait (111111) soit au pire état décrit par le SF6Dv2 (555655). Chaque résultat du traitement est associé à une fréquence d'occurrence (c.-à-d. une probabilité de succès) qui se modifie selon la procédure du Pingpong développée par une équipe de l'Université McMaster en Ontario (Furlong et al 1990; Torrance et Feeny 1989). Et la méthode du SG consiste à varier cette distribution de probabilité jusqu'à ce que le répondant demeure indifférent non pas en univers incertains qui renferment uniquement les résultats probabilistes, mais entre l'état jugé imparfait qui est certain de se réaliser (alternative 1) et le traitement risqué (alternative 2). Au point d'indifférence, la probabilité associée à l'état parfait (à l'intérieur du traitement) représente l'utilité de l'état présent dans l'alternative 1. Lorsque ce score d'utilité est multiplié par la durée (en année) il est égal au QALY sur ce nombre d'années. Puisque dans cette étude les QALY sont considérés sur une durée d'une année, le score d'utilité au point d'indifférence sera considéré comme un QALY.

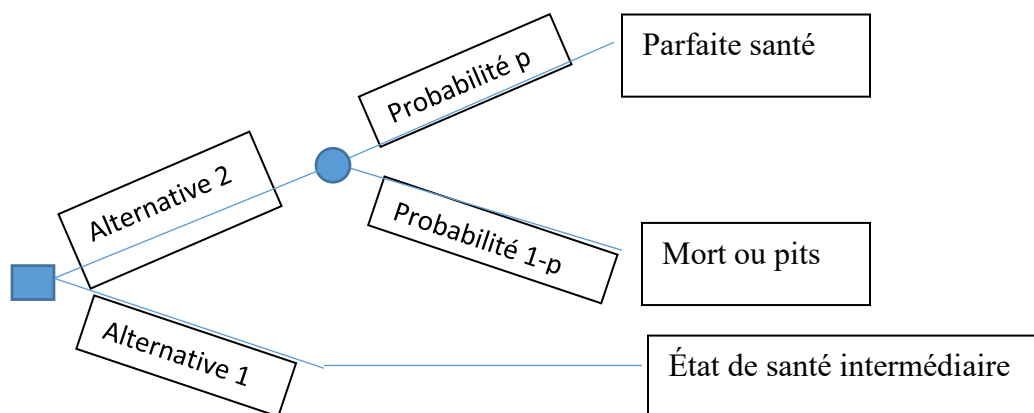


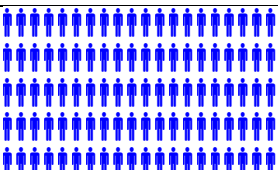
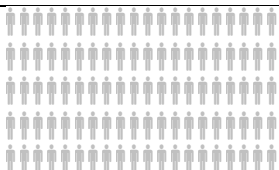


Figure 1 : description de l'évaluation d'un état de santé à travers le SG

Tout comme Brazier et al (2002) nous allouons un ensemble de 5 états de santé intermédiaires pour chaque individu de la façon précédemment décrite. Voici un exemple tiré du questionnaire SG:

« **Le premier état de santé dans une série de 5** pour lequel on vous demande de procéder à un choix est décrit par la carte suivante :

<b>Choix A :</b> État de santé <b>certain à 100%</b> <b>100 chances sur 100</b> <i>Carte 113411</i>		
<b>Vous vivez pendant 10 ans dans l'état de santé suivant puis vous mourrez :</b>		
<b>Pas du tout limité</b> dans les activités exigeant un <b>effort physique important</b>		
<b>Jamais accomplir</b> <b>moins de choses</b> que vous l'auriez voulu dans vos activités quotidiennes		
<b>Activités sociales</b> parfois limitées		
<b>Douleurs moyennes</b>		
<b>Jamais très nerveux ou triste et démoralisé</b>		
<b>Jamais épuisé et vidé</b>		

Figure 2 : description d'un état de santé quelconque  
 Afin d'améliorer votre état de santé tel que décrit ci-dessus, un traitement est disponible.

<b>Choix B :</b> 100% de chance de retrouver un état de santé parfait et 0% de chance de tomber dans le pire état de santé possible.			
<b>100% de chance de succès</b> <b>100 chances sur 100</b> <i>Carte 111111</i>		<b>0% de chance d'échec</b> <b>0 chance sur 100</b> <i>Carte 555655</i>	
			
<b>Vous vivez pendant 10 ans dans l'état de santé suivant puis vous mourrez :</b>		<b>Vous vivez pendant 10 ans dans l'état de santé suivant puis vous mourrez :</b>	
<b>Pas du tout limité</b> dans les activités exigeant un <b>effort</b>		<b>Très limité pour prendre un bain ou m'habiller</b>	

<b>physique important</b>			
<b>Jamais accomplir moins de choses</b> que vous l'auriez voulu dans vos activités quotidiennes		<b>Tout le temps accomplir moins de choses</b> que vous l'auriez voulu dans vos activités quotidiennes	
<b>Activités sociales jamais limitées</b>		<b>Activités sociales tout le temps limitées</b>	
<b>Aucune douleur</b>		<b>Douleurs très intenses</b>	
<b>Jamais très nerveux ou triste et démoralisé</b>		<b>Tout le temps très nerveux ou triste et démoralisé</b>	
<b>Jamais épuisé et vidé</b>		<b>Tout le temps épuisé et vidé</b>	

Figure 3 : présentation des résultats du traitement dans le SG  
1.1. Quelle situation préférez-vous ?

Choix A ☐ (aller au 1.13)  
Choix B (avec traitement) ☐ (aller au 1.2)  
Je n'arrive pas à choisir ☐ (aller au 1.15) »

On modifie la probabilité jusqu'à ce que l'individu soit indifférent entre l'état du choix A et le traitement B.

Nous remplacerons dans un 6<sup>ém</sup> exercice de SG l'état de santé intermédiaire par le pits (555655). Et suivant la réponse qu'il aura donnée à la question de savoir s'il préfère ou non la mort au pits, on assignera des valeurs négatives ou positives respectivement à ce PITS.

Les probabilités sont assignées aux états de santé conformément à l'arbre des valeurs décrit en annexe 1. A1 représente le même état de santé tout au long du processus et issu de la sélection orthogonale décrite précédemment. Gardons en tête que c'est l'état de santé auquel on veut déterminer le QALY. A2 est le même traitement qui est proposé avec différente chance de succès.

Ainsi il transforme chaque décision prise en un processus itératif jusqu'à l'obtention du point d'indifférence. Comme on peut le constater, le répondant n'est pas obligé de terminer le processus de Pingpong. Dès que la probabilité d'indifférence (c.-à-d. QALY) pour un état est trouvée, le processus s'arrête et il bascule immédiatement au second état jusqu'à l'épuisement des 5.







Pour le DCE : les 120 états sont répartis en 12 blocs de 10 paires. Les blocs sont répartis de façon aléatoire entre les répondants.

Le choix expérimental discret se décrit de la façon suivante : il présente deux états de santé et demande aux répondants de choisir l'un d'entre eux sur la base de leur préférence. L'état qui est choisi reçoit la valeur de 1 et l'autre une valeur nulle. À ce stade on n'a pas de QALY, mais retenons que le choix est transformé en une variable dichotomique pour les besoins du modèle de régression. Voici un exemple tiré de la partie DCE de l'enquête:

« Dans les questions suivantes, nous vous présentons deux situations de santé fictives qui durent un certain nombre d'années.

Nous voulons que vous nous disiez dans lequel des deux états de santé décrits vous préféreriez vivre

1.1. Voici la première carte de choix

Description de santé A		Description de santé B	
Carte 453215		Carte 224612	
Vous vivez dans l'état de santé suivant puis vous mourrez :		Vous vivez dans l'état de santé suivant puis vous mourrez :	
6 années de vie		10 années de vie	
Très limité dans les activités exigeant un effort physique modéré		Un peu limité dans les activités exigeant un effort physique important	
Tout le temps accomplir moins de choses que vous		Rarement accomplir moins de choses que vous l'auriez voulu	

l'auriez voulu dans vos activités quotidiennes		dans vos activités quotidiennes	
<b>Activités sociales parfois limitées</b>		<b>Activités sociales la plupart du temps limitées</b>	
<b>Douleurs très légères</b>		<b>Douleurs très intenses</b>	
<b>Jamais très nerveux ou triste et démoralisé</b>		<b>Jamais très nerveux ou triste et démoralisé</b>	
<b>Tout le temps épuisé et vidé</b>		<b>Rarement épuisé et vidé</b>	

Figure 4 : confrontation des états de santé dans le DCE

Quelle situation préférez-vous ?

Description de santé A

☐ (aller à la question 1.2)

Description de santé B

☐ (aller à la question 1.3) »

Ce procédé sera répété 10 fois pour chaque répondant. À noter que la construction de ces paires d'états de santé a été réalisée avec le logiciel économétrique SAS qui s'est assuré de ne pas inclure la dominance entre les états. Ainsi des deux états confrontés dans chaque paire, aucun ne domine l'autre. On s'est aussi assuré à ce qu'aucune paire ni aucun état de santé ne soient répétés au moment de la création du design orthogonal dans le logiciel SAS (c.-à-d. aucun état de santé ne se retrouve dans plus d'une paire).

### 3. Les données

Au total, 1958+726 personnes ont été approchées, dont 1958 avec la version de questionnaire DCE et 726 avec le questionnaire SG. Ce décalage est dû au fait que l'on dispose d'une seule version du SG contre 3 pour le DCE dans la base de données utilisée. Avec un taux de réponse (qui est ici défini comme l'ensemble des individus qui ont répondu à au moins une question dans les deux exercices) de 60% pour SG et 68% pour DCE, nous avons au final 1761 individus, dont 1325 associés au DCE et 436 associés au SG.

Parmi les 436 individus ayant été alloués à un groupe de SG, 60 ont été exclus pour l'unique raison qu'ils ont associé le même QALY aux 5 états qui leur ont été alloués. À ce niveau c'est l'individu qui est exclu.

Tableau 1. Caractéristique des individus ayant participé à l'épreuve du SG

	Inclus	Exclus	Différence moyenne standardisée
Observations	376	60	
Âge	49 (15)	46 (13)	-0,2
IMC	28 (8)	29 (6)	0,1
Revenu minimum	51795	54350	0,08
Revenu maximum	63403	66415	0,08
Zone rurale	130 (0,35)	20 (0,33)	-0,02
Zone urbaine	246 (0,65)	40 (0,67)	
Propriétaire	224 (0,6)	23 (0,38)	0,04
Locataire	152 (0,4)	37 (0,62)	
Employé autonome	195 (0,52)	36 (0,6)	-0,16
Sans emploi	18 (0,05)	3 (0,05)	
À la maison	26 (0,07)	4 (0,06)	
Retraité	92 (0,24)	12 (0,2)	
Congé maladie	10 (0,03)	0	
Étudiant	22 (0,06)	3 (0,05)	
Autre	13 (0,03)	2 (0,03)	
Maladie grave	81 (0,21)	14 (0,23)	0,04
Pas de maladie grave	295 (0,79)	46 (0,77)	
Diplôme professionnel	303 (0,81)	52 (0,87)	0,16

Pas de diplôme	73 (0,19)	8 (0,13)	
Homme	171 (0,45)	25 (0,42)	0,08
Femme	205 (0,55)	35 (0,58)	
Exercice facile	121 (0,34)	24 (0,4)	0,28
Exercice difficile	140 (0,43)	21 (0,35)	

Les nombres entre parenthèses sont des variances pour les variables continues et des fréquences relatives pour les variables catégoriques. Les caractéristiques des individus inclus et exclus sont données dans le tableau ci-dessus. Les 60 répondants exclus sont des individus qui sont en moyenne plus jeune, mais présentent un surpoids plus élevé par rapport à ceux qui sont inclus. Ils ont en moyenne un revenu légèrement au-dessus des inclus. Ils se localisent généralement en zone urbaine et regroupent plus d'individus qui rapportent être sans emploi. C'est dans ce groupe d'exclus que l'on trouve plus d'individus qui ont des problèmes liés à la santé. Ils semblent mieux instruits, sont généralement des femmes et prétendent que l'épreuve du SG est facile. En utilisant des questions relatives à la compréhension, Ratcliffe et al. (2011) ont trouvé que seulement 1/4 des individus qui disent comprendre le SG sont capables de fournir des réponses utilisables. Même si son étude est basée sur un échantillon composé uniquement de jeunes adolescents, il semble que l'on ait des résultats similaires dans des proportions différentes dans la mesure où 40% de ceux qui sont exclus trouvent la tâche du SG facile. La dernière colonne présente la différence moyenne standardisée entre les deux groupes. Elle est calculée en faisant la différence entre la moyenne des individus inclus et exclus que l'on rapporte à l'écart-type de l'échantillon. Cet indicateur montre une absence de différence statistiquement significative entre les individus inclus et exclus mis à part la dernière variable « Exercice » où la différence demeure significative, mais faible.

Un autre nettoyage est fait, mais cette fois au niveau des QALY associés aux différents états de santé. En effet, pour certains individus, la valeur de QALY qui est associée à l'un des états alloués peut être jugée incohérente. Par exemple, nous avons dû supprimer 137 réponses où le répondant choisit le traitement dont les chances de succès sont nulles, ce qui



représente clairement un problème de compréhension. Cela équivaut à dire que l'état de santé que l'on évalue (dans l'alternative 1) est pire que le PITS, chose qui est impossible. Dans le 6<sup>e</sup> exercice du SG où l'on demandait une évaluation du PITS, 69 réponses ont également été supprimées. Par exemple, lorsque le répondant dit préférer la mort au PITS mais choisit un traitement le menant au PITS alors qu'il peut mourir immédiatement et sans douleur. Pour ces deux derniers cas, on ne peut pas supprimer l'individu, mais uniquement la réponse qui est incohérente dans la mesure où le concerné peut fournir d'autres réponses qui sont cohérentes dans l'évaluation d'autres états de santé. Rappelons que chaque individu évalue un total de 5 états de santé et le PITS. Ainsi, il peut fournir une ou deux réponses non cohérentes et d'autres réponses jugées cohérentes que l'on risque de perdre si on exclut l'individu. Les 376 individus inclus devraient fournir 2256 valeurs de QALY pour les l'évaluation des 61 états de santé évalués. Si l'on soustrait les évaluations incohérentes, on se retrouve avec 2050 (2256-137-69) valeurs de QALY. De ces 2050 évaluations, 216 sont absentes. Ce qui nous ramène à un total de 1834 valeurs de QALY dont les caractéristiques sont résumées dans la section suivante.

La description que l'on apporte dans cette partie concerne uniquement une évaluation des états de santé par la méthode du SG. En effet, on n'a pas trouvé une distinction particulière à apporter par rapport aux états évalués par le DCE puisque ce dernier demande un simple classement sur la base des préférences. Le tableau suivant résume les caractéristiques des 61 états concernés.

Tableau 2. Descriptions statistiques de l'évaluation (QALY) des 61 états de santé du SF6Dv2

États	N	min	max	Médiane	Moyenne	Variance	Quartile(25)	Quartile(75)	Skewness
111133	21	-0,5	1	0,75	0,61	0,41	0,1	0,975	-0,48
111245	21	0,05	1	0,65	0,6	0,35	0,35	0,975	-0,13
111415	28	0,05	1	0,6	0,52	0,36	0,125	0,85	-0,03
112333	33	0,05	1	0,65	0,65	0,34	0,35	1	-0,43
123343	24	-0,05	1	0,625	0,59	0,36	0,2	0,9625	-0,22
134411	29	0,05	1	0,8	0,7	0,29	0,45	0,975	-0,74
135334	38	0,05	1	0,65	0,55	0,36	0,15	0,925	-0,23
143315	26	-0,317	1	0,45	0,48	0,38	0,15	0,85	-0,07
143432	23	-0,05	1	0,65	0,62	0,32	0,35	0,925	-0,47

145423	35	-0,05	1	0,85	0,62	0,37	0,15	0,975	-0,57
211244	20	-0,05	1	0,8625	0,65	0,4	0,25	1	-0,5
213232	36	-0,05	1	0,75	0,63	0,35	0,45	0,975	-0,62
213411	28	0,05	1	0,6	0,6	0,34	0,35	0,95	-0,29
214322	32	0,05	1	0,8	0,64	0,35	0,375	0,925	-0,69
222444	28	-0,3	1	0,725	0,64	0,36	0,4	0,95	-0,92
232514	28	0,05	1	0,65	0,57	0,37	0,15	0,975	-0,24
235342	33	0,05	1	0,6	0,56	0,37	0,15	0,975	-0,15
241423	35	-0,15	1	0,75	0,63	0,37	0,25	0,975	-0,67
244225	30	-0,65	1	0,4	0,38	0,4	0,1	0,7	-0,3
244231	23	0,05	1	0,8	0,68	0,35	0,45	1	-0,76
253253	19	0,05	1	0,45	0,5	0,36	0,15	0,975	0,25
254554	34	-0,45	1	0,45	0,44	0,44	0,05	0,975	0,07
255621	25	0,05	1	0,55	0,53	0,34	0,25	0,85	0,02
314252	28	0,05	1	0,65	0,57	0,38	0,15	0,9875	-0,15
314321	25	-0,32	1	0,4	0,47	0,44	0,05	0,925	-0,05
315455	24	0,05	1	0,7	0,62	0,34	0,35	0,9625	-0,37
324521	26	0,05	1	0,65	0,53	0,35	0,1	0,65	-0,28
325435	29	-0,317	1	0,35	0,47	0,38	0,15	0,85	0,01
331131	32	0,05	1	0,65	0,63	0,33	0,375	0,975	-0,41
333614	29	0,05	1	0,45	0,5	0,34	0,25	0,85	0,25
334124	18	0,05	1	0,4	0,5	0,38	0,15	0,975	0,27
334443	29	0,05	1	0,45	0,51	0,38	0,15	0,925	0,05
341334	33	0,05	1	0,5	0,5	0,32	0,25	0,85	0,05
352211	33	0,05	1	0,65	0,64	0,36	0,35	0,975	-0,57
353345	30	0,05	1	0,425	0,45	0,33	0,1	0,7	0,2
354145	27	0,05	1	0,65	0,64	0,36	0,3	1	-0,37
355522	27	0,05	1	0,45	0,52	0,37	0,05	0,975	0,03
413413	24	-0,15	1	0,65	0,58	0,38	0,25	1	-0,36
414442	27	-0,31	1	0,75	0,64	0,4	0,5	1	-0,93
423342	28	0,05	1	0,5	0,5	0,31	0,2	0,825	0,11
434113	27	0,05	1	0,75	0,6	0,36	0,15	0,975	-0,42
435324	24	0,05	1	0,6	0,56	0,28	0,4	0,75	-0,4
442625	26	0,05	1	0,65	0,54	0,42	0,05	1	-0,06
443614	20	-0,8	1	0,4	0,44	0,48	0,1	1	-0,56
452625	26	-0,25	1	0,15	0,33	0,36	0,05	0,65	0,66
454343	24	-0,317	1	0,65	0,59	0,37	0,375	0,925	-0,75
511141	22	0,05	1	0,65	0,64	0,31	0,45	0,975	-0,37
512143	26	-0,3	1	0,45	0,47	0,42	0,1	0,975	0,07
515224	21	-0,32	1	0,55	0,51	0,39	0,25	0,75	-0,35
531622	26	-0,05	1	0,45	0,46	0,35	0,1	0,75	0,13

533451	30	-0,31	1	0,375	0,47	0,45	0,05	1	-0,04
535645	22	-0,05	1	0,3	0,47	0,4	0,3	0,9	0,19
541214	28	0,05	1	0,75	0,65	0,34	0,4	0,975	-0,69
542132	24	0,05	1	0,625	0,6	0,33	0,3	0,8875	-0,35
542315	20	0,05	1	0,55	0,51	0,42	0,1	0,975	0,08
543341	29	-0,15	1	0,25	0,35	0,33	0,05	0,65	0,51
552511	28	0,05	1	0,65	0,57	0,38	0,15	0,9875	-0,2
552623	31	-0,8	1	0,45	0,47	0,45	0,1	1	-0,45
553222	21	0,05	1	0,65	0,4	0,37	0,05	0,65	0,55
554542	24	-0,05	1	0,275	0,46	0,43	0,05	1	0,28
555655	217	-0,975	1	-0,25	-0,32	0,3	-0,55	-0,05	-0,33

Tous les états ont été évalués au moins 20 fois à l'exception de deux (18 et 19 fois). Si l'on ne prend pas en compte les valeurs négatives, on voit que le QALY est compris entre 0 et 1 (colonnes min max). La présence de valeurs négatives sur la colonne des minimas s'explique du fait que certains des individus de l'échantillon ont répondu " *je n'arrive pas à choisir* " entre un des états qui leur a été proposé et le traitement lorsque celui-ci apparaît avec une chance de succès nulle. Ce comportement laisse entendre que l'individu associe une utilité identique entre l'état de santé évalué et le PITS. Précisons que, contrairement au groupe de 137 réponses que l'on a enlevées de l'échantillon précédemment (dont le comportement montre une préférence du PITS au détriment d'un état intermédiaire), les réponses auxquelles nous nous référons ici sont celles où l'individu associe une utilité similaire entre l'état de santé évalué et le PITS. Dans ces cas de figure, on va chercher la valeur du pire état pour cet individu et venir l'associer à l'état dont le répondant rapporte être indifférent à ce PITS. Deux cas de figure sont possibles dans cette situation :

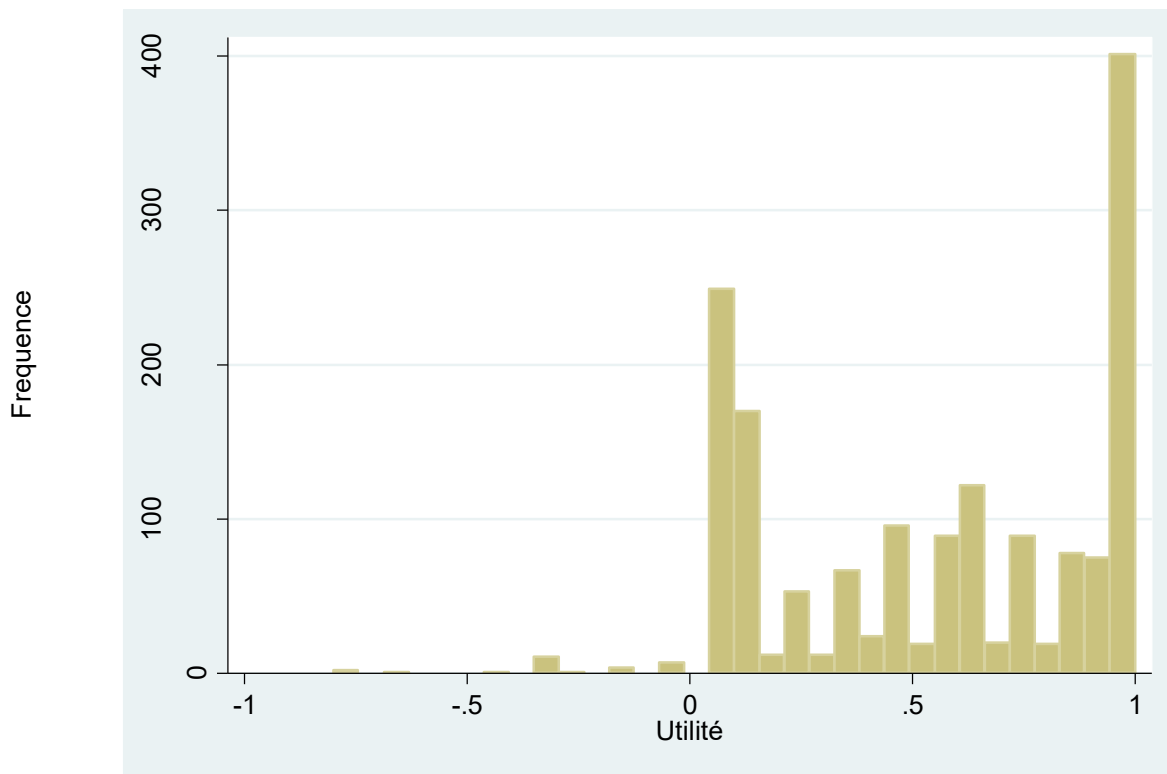
Cas 1 : soit il préfère le pire état à la mort

Cas 2 : la mort est préférée au pire état.

Cette étape nous permet de savoir la façon dont le SG sera libellé pour chaque répondant suivant qu'il se trouve dans le cas 1 ou le cas 2.

Sachant que la mort est évaluée à zéro, on obtient des QALY du PITS qui sont négatifs dans le cas 2. Ce qui explique la présence de valeurs négatives dans le tableau précédent. Environ 60% de l'échantillon SG a répondu à cette 6<sup>ém</sup> question (217 sur 376) dont le but

est d'évaluer le PITS. Au total, 95% de ceux qui ont répondu préfèrent mourir plutôt que de vivre dans le pire état pendant 10 ans (206 sur 217). À la 5<sup>ème</sup> colonne du tableau 2, on peut observer que les états de santé sont évalués à une moyenne située entre 0,4 et 0,6 avec une forte variance qui peut aller jusqu'à 0,48. Tous ces 60 états présents ont été évalués au moins une fois comme étant égale à 1 (santé parfaite, colonne Max) laissant entrevoir soit un échantillon qui présente une aversion au risque soit un échantillon qui éprouve de la difficulté à comprendre le SG. Ce qui est observable dans l'histogramme ci-dessus avec un pic au niveau du 1.



OBS	1622
Moyenne	0,45
Variance	0,46
Skewness	-0,52
kurtosis	2,6
min	-0,975
Max	1

Figure 5 : Histogramme descriptif de l'évaluation des états de santé

L'histogramme ci-dessus ne présente pas d'informations qui se rattachent au PITS par peur d'une double comptabilisation de certains QALY associés à la fois au PITS et à certains états lorsque le répondant dit être indifférent entre le traitement qui a des chances de succès nulles et l'alternative 1. Ainsi, il représente l'histogramme de 1622 (1839-217) valeurs de QALY provenant de l'évaluation des 60 états de santé. L'ordonnée

à l'origine représente une fréquence, c'est-à-dire le nombre de fois qu'une valeur de QALY apparaît lors de l'évaluation. Sur la ligne des abscisses, on a les différentes valeurs d'utilités ou QALY.

Tout comme (Brazier et al. 2002), nous observons des données qui présentent une forte asymétrie à l'échelle d'évaluation des états. Des valeurs négatives sont présentes, mais à une fréquence très faible (27/1622). Finalement, des 376 individus inclus, 217 ont évalué le PITS. De ce nombre, 206 individus préfèrent mourir immédiatement plutôt que de vivre dans le pire état de santé et seulement 27 individus ont manifesté de l'indifférence entre un des états alloués et le traitement lorsque celui-ci apparaît avec des chances de succès nulles. Les données décrites ci-dessus sont caractérisées par une asymétrie, mais aussi une distribution anormale. Les valeurs de QALY associé aux différents états de santé sont diverses et variées. Sachant que les états ont été répartis entre les individus de façon aléatoire, cette différence reflète la diversité des préférences des individus en matière de santé et se situe entre les individus.

Pour le DCE, il est justifié de considérer l'étape antérieure de classement des 6 états de santé comme un outil permettant de contrôler le niveau de compréhensions des individus dans la mesure où il demande un simple classement entre alternatives. Cette étape va nous permettre de distinguer les répondants selon leur niveau de compréhension et constituer ainsi un outil de nettoyage des données. Un autre moyen est de déceler ceux qui choisissent à chaque fois la première alternative (respectivement la deuxième) et de les supprimer. Des 1325 individus qui ont fini l'épreuve du DCE, 276 sont retirés du groupe pour des raisons d'incompréhensions. On retrouve dans ce groupe de 276 en général des individus qui disent préférer le pire état à un état intermédiaire ou même la santé parfaite. Ce qui nous ramène à un total de 1049 auquel on retire 120 autres du fait qu'ils n'ont pas modifié leur choix sur une série de 10 paires d'états de santé qu'on leur a présentée (toujours choix A ou toujours choix B). Remarquons que les raisons qui nous ont poussés à éliminer ces individus dans cet exercice de DCE sont exactement les mêmes que l'on retrouve dans le SG. La seule différence est que pour le SG, on les rencontre au fur et à mesure que l'on évolue, dans la procédure. On se retrouve finalement avec un total de 929 individus. Les caractéristiques de tous ces individus inclus et exclus sont résumées dans le tableau suivant.

Tableau 3. Caractéristique des individus ayant participé à l'épreuve du DCE

	Inclus	Exclus	Différence moyenne standardisée
Observations	929	396	
Âge	47 (14)	49 (13)	0,13
IMC	27 (7)	28 (6)	0,05
Revenu minimum	52976	47512	-0,16
Revenu maximum	65144	58471	-0,16
Zone rurale	295 (0,32)	139 (0,35)	0,07
Zone urbaine	634 (0,68)	257 (0,65)	
Propriétaire	552 (0,59)	205 (0,52)	-0,15
Locataire	377 (0,41)	191 (0,48)	
Employé/ autono	455 (0,49)	196 (0,49)	-0,1
Sans emploi	54 (0,06)	29 (0,07)	
À la maison	57 (0,06)	31 (0,07)	
Retraité	230 (0,25)	109 (0,28)	
Congé maladie	25 (0,02)	6 (0,01)	
Étudiant	79 (0,08)	13 (0,03)	
Autre	29 (0,03)	12 (0,03)	
Maladie grave	190 (0,2)	101 (0,26)	0,12
Pas de maladie grave	739 (0,8)	295 (0,74)	
Diplôme professionnel	776 (0,84)	321 (0,81)	-0,06
Pas de diplôme	153 (0,16)	75 (0,19)	
Homme	352 (0,38)	182 (0,46)	-0,16
Femme	577 (0,62)	214 (0,54)	
Exercice facile	275 (0,32)	123 (0,35)	0,08

Exercice difficile	332 (0,39)	131 (0,37)
-----------------------	------------	------------

Les nombres entre parenthèses sont des variances pour les variables continues et des fréquences relatives pour les variables catégoriques. Le tableau ci-dessus présente les caractéristiques des individus qui sont à la fois inclus et exclus par le DCE. On peut constater que les individus exclus sont plus vieux et ont un IMC qui reflète un surpoids plus élevé par rapport aux inclus. Ils ont des revenus plus faibles. Généralement localisés dans les zones rurales où ils sont en location. Comme dans le SG, ce groupe enregistre plus d'individus qui rapportent être sans emploi et dispose d'un problème lié à leur propre état de santé. Dans la mesure où ils sont locataires, ils auront plus de difficultés à faire face à certaines dépenses comme le loyer. C'est aussi dans ce groupe que l'on retrouve un plus grand nombre d'individus de sexe masculin et qui jugent l'épreuve légèrement moins facile contrairement au SG. Comme dans le SG, notons que la différence standardisée révèle une différence très réduite, voire inexistante, entre les deux groupes.

Si on considère les individus inclus, on peut voir que dans les deux méthodes, ils jugent l'épreuve plutôt difficile (voir la variable *difficulté exercice* dans les tableaux 1 et 3). Finalement, on se retrouve avec 929 individus qui évaluent 10 paires d'états de santé chacun pour un total de 18580 ( $929 \times 20$ ) évaluations. Mais notons que 315 paires n'ont pas reçu d'évaluations soit un échantillon final de 17380 ( $18580 - 630$ ).

Les QALY associées aux différents états de santé pour les deux méthodes sont variées et comprises entre 0 et 1 si on exclut les 27 évaluations des individus indifférents au PITS.

La différence du nombre d'individus entre les deux méthodes est due au fait que dans le DCE on dispose de 3 versions (durée, probabilité et mort subite) qui sont toutes utilisables dans cette étude contre une seule pour le SG.

#### 4. Le modèle

Le but de chaque méthode est de fournir le meilleur modèle de prévision de l'ensemble des états de santé que renferme le SF6Dv2. En d'autres termes, on cherche à prédire la valeur de QALY des différents états de santé en fonction des niveaux de chaque dimension. Rappelons que chaque état de santé est défini selon ces différents niveaux des 6

dimensions. Ainsi, avoir une bonne estimation de l'utilité de ces derniers nous permet de prédire un QALY plus exact pour l'ensemble des états de santé que présente le SF6Dv2. On s'attend à avoir autant d'observations que le nombre de fois que les états ont été évalués par les répondants, que ce soit du côté du SG (1834) ou du DCE (17320).

La forme générale du modèle est donnée à l'équation 1. Cependant, plusieurs spécifications peuvent être utilisées suivant l'hypothèse que l'on fait sur la distribution du terme d'erreur ou la nature de la variable expliquée :

$$Y_{ij} = (\beta' A_{lm} + \lambda' B)_{ij} + \mu_{ij}(1)$$

$Y_{ij}$  : QALY de l'état  $j$  ( $j = 1 \dots 60^2$  pour le SG et  $1 \dots 120$  pour le DCE) évalué par l'individu  $i$  ( $i = 1 \dots 376$  pour le SG et  $1 \dots 929$  pour DCE).  $\beta$  est le vecteur des coefficients associé à chaque niveau des 6 différentes dimensions du SF6Dv2.  $A_{lm}$  est la matrice des variables explicatives (niveaux dans chaque dimension) ( $l = 1 \dots 6$  et  $m = 1 \dots 6$ ). Par exemple, le  $A_{52}$  représente le niveau 2 de la dimension 5 soit *rarement très nerveux ou triste et démoralisé*.

Chaque niveau est par la suite transformé en variable muette avant d'être inclus dans la régression.

$A_{52} = 1$  si à la dimension santé mentale le niveau est de 2 dans cet état

$A_{52} = 0$  si à la dimension santé mentale le niveau n'est pas égal à 2 dans cet état

Le SF6Dv2 se résume à un total de 31 niveaux repartis dans les 6 différentes dimensions. Le premier niveau de chaque dimension est omis de la régression pour éviter le problème de la colinéarité parfaite. De plus si l'on réunit ces niveaux omis, on se retrouve avec la santé parfaite dont le QALY est égal à 1.

$$Y_{ij} = \beta_1 PF_2 + \beta_2 PF_3 + \beta_3 PF_4 + \beta_4 PF_5 + \beta_5 RL_2 + \beta_6 RL_3 + \beta_7 RL_4 + \beta_8 RL_5 + \beta_9 SF_2 + \beta_{10} SF_3 + \beta_{11} SF_4 + \beta_{12} SF_5 + \beta_{13} PAIN_2 + \beta_{14} PAIN_3 + \beta_{15} PAIN_4 + \beta_{16} PAIN_5 + \beta_{17} PAIN_6 + \beta_{18} MH_2 +$$

<sup>2</sup> On se retrouve avec 60 états de santé au lieu de 120 dans le SG à cause d'un bug de programmation dans le questionnaire en ligne.



$$\beta_{19}MH_3 + \beta_{20}MH_4 + \beta_{21}MH_5 + \beta_{22}VIT_2 + \beta_{23}VIT_3 + \beta_{24}VIT_4 + \beta_{25}VIT_5 + \lambda_1 MOST_{ij} + \lambda_1 LEAST_{ij} + \mu_{ij} \quad (2)$$

Les termes d'interaction sont représentés dans l'équation 1 par la variable  $B_{ij}$  par les variables MOST et LEAST dans l'équation 2. *On fait ici le choix d'adopter la même méthode que (Brazier et al. (2002) du fait qu'il existe une multitude de termes d'interaction et que l'on ne dispose pas de moyens nous permettant de contrôler l'hypothèse d'indépendance entre eux ou même de choisir ceux qui auraient plus d'impact dans la régression.* Ce vecteur est défini comme celui des valeurs extrêmes. Rappelons que le modèle ci-dessus est réalisé à l'échelle des états de santé. Chaque niveau de chaque état de santé (variable explicative de l'équation 2) est transformé en variable muette et prend la valeur de 1 ou de 0. Lorsqu'un niveau est présent dans un état il prend la valeur de 1 et 0 sinon. Les variables MOST et LEAST sont définies de la même façon, c'est à dire en variables muettes sauf qu'ils prennent la valeur de 1 lorsque l'état se présente avec des niveaux 1 ou 2 pour LEAST (moins sévère), des niveaux 4, 5 ou 6 pour MOST (plus sévère) dans chaque dimension et 0 sinon. Tout comme (Brazier et al. 2002), nous faisons le choix de présenter les résultats avec et sans les termes d'interactions pour mieux observer leur impact sur l'estimation de la valeur du QALY.

$\mu_{ij}$  représente l'erreur de mesure dépendante de la spécification que l'on donne à la fonction de régression. Nous optons pour le modèle additif qui n'impose aucune restriction quant à l'ordre des niveaux. La fonction de régression prend dans un premier temps la forme des moindres carrés ordinaires qui est une projection orthogonale tout en minimisant les erreurs de mesures.

$$\mu_{ij} = e_j + \varepsilon_{ij} \quad (3)$$

Le terme  $e_j$  représente le terme d'erreur de chaque répondant supposé aléatoire entre les individus.  $\varepsilon_{ij}$  est le terme d'erreur qui provient de l'évaluation de l'état de santé  $j$  par l'individu  $i$  et qui est supposé aléatoire entre les observations.  $Cov(e_j \varepsilon_{ij}) = 0$  ce qui signifie que les états ont été répartis entre les individus de façon aléatoire.

Si  $Cov(A_{lm}, e_j)$  est nulle on utilise le modèle à effet aléatoire, par contre si cette hypothèse n'est pas respectée le modèle à effet fixe est recommandé.

Le test de Hausman déterminera par la suite l'adéquation d'une spécification à effet fixe ou à effet aléatoire et enfin l'application de certains modèles agrégés par la moyenne et la médiane qui respecte les mêmes hypothèses que les moindres carrés ordinaires.

Ces derniers sont couramment utilisés dans les études de QALY de Brazier et al. (2002, 2009). Ce sont des modèles linéaires tirés de la minimisation des erreurs de mesures. L'idée est de trouver si une approche par valeur moyenne peut avoir plus de signification en matière de QALY puisque les préférences des individus sont assez distinctes. L'utilité d'un état de santé est donnée dans ce cas par la moyenne des utilités qui sont fournies par différents individus à un état de santé et qui constituent pour ce modèle la variable expliquée. Ces modèles agrégés gardent une spécification identique à l'équation 2 à l'exception de l'indice  $j$  qui est absent à cause du calcul de la moyenne et de la médiane. Puisque ce sont des modèles agrégés, on s'attend à avoir une erreur absolue moyenne plus faible associée à un plus grand pouvoir explicatif.

À noter que pour chaque modèle (effet fixe et agrégé), on a proposé une réplique qui garde les mêmes caractéristiques mis à part la constante qui est contrainte à l'unité. L'idée sous-jacente est d'essayer de se rapprocher le plus possible de la procédure de calcul du QALY. Ce dernier est obtenu en déduisant l'utilité de chaque niveau qui définit un état de santé de la constante. Fixer la constante à l'unité revient à dire que les QALY des états de santé sont déduits à partir de la santé parfaite dont le QALY est évalué à 1.

Mais il arrive parfois que notre variable dépendante ne soit pas continue, comme c'est le cas pour le DCE où la variable dépendante prend la valeur 0/1. On a dans ce cas une variable dépendante binaire (0/1). Si  $\mu_{ij}$  respecte l'hypothèse de l'espérance conditionnelle stricte nulle on peut estimer ce modèle par les MCO. Mais dans ce cas, la variable  $Y_{ij}$  peut prendre des valeurs au-delà de l'intervalle 0 et 1 et le terme d'erreur risque d'être hétéroscédastique. Pour faire face à ces problèmes, on utilise deux alternatives :

## Le LOGIT conditionnel

Le modèle des choix discrets est basé sur le comportement de maximisation de l'utilité de l'individu. Un individu choisit l'alternative  $j$  parmi les alternatives  $k$  si l'utilité associée à l'alternative  $i$  est supérieure.

Typiquement, dans les modèles de logit conditionnel, l'utilité de l'individu se décrit de la manière suivante :

$$U_{ij} = X_{ij}\beta + \varepsilon_{ij} \quad (4)$$

$U_{ij}$  est le choix  $j$  de l'individu  $i$ .  $X_{ij}$  étant les variables observables.  $\varepsilon_{ij}$  est le terme d'erreur suivant une distribution *iid* avec valeur extrême type I (loi Gamble). Le modèle suppose que les paramètres estimés  $\beta$  ne varient pas entre les individus. Cela suppose une homogénéité des préférences.

Une des grosses limitations des modèles de Logit conditionnel est l'hypothèse appelée IIA (Independence of Irrelevant Alternatives). Cette restriction suppose qu'il n'existe pas de corrélations des termes d'erreurs entre les alternatives  $i$  et  $k$ . C'est McFadden (1973) qui a introduit ce modèle. La régression logistique conditionnelle est basée sur un modèle proche de celui de la régression logistique. La différence vient de ce que tous les individus sont soumis à différentes situations avant d'exprimer leur choix (sous forme d'une variable binaire qui constitue la variable dépendante). Le modèle prend en compte la distinction entre les individus et le nombre de CHOICE SET évaluée par chaque individu.

## Le modèle Logit à coefficient aléatoire

Le modèle Logit à coefficient aléatoire relâche la condition IIA. Contrairement au logit conditionnel, le modèle à coefficient aléatoire permet que les paramètres estimés varient entre les individus. En d'autres termes, ce modèle autorise une hétérogénéité des préférences :

$$U_{ij} = X_{ij}\beta_i + \varepsilon_{ij} \quad (5)$$

Dans cette équation  $U_{ij}$  représente le choix  $j$  de l'individu  $i$ . Les  $\beta_i$  varient entre les individus et pourraient suivre une distribution normale, log normal, uniforme, etc.

Le modèle logit à coefficients aléatoires est le modèle le plus prometteur parmi les modèles de choix discret (C. Bonnet 2001). C'est un modèle qui se base sur la lignée des modèles logit.

Après avoir obtenu les résultats de chaque régression, nous utiliserons des tests et des critères de choix pour déterminer la meilleure spécification capable de donner la meilleure prévision du QALY, d'abord entre les différentes spécifications d'une méthode et entre les deux méthodes. Ces critères de choix sont donnés ci-dessous :

- 1) Le nombre de paramètres significatifs,
- 2) Le nombre d'incohérences (signe non négatif et taille du coefficient décroissante),
- 3) Le pouvoir explicatif du modèle ( $R^2$ )
- 4) Le critère d'information d'Akaike,
- 5) L'erreur absolue moyenne qui est calculée en faisant la moyenne de la valeur absolue de chaque différence entre la valeur réelle et la valeur prédite,
- 6) Le test de JB pour vérifier la distribution des erreurs de prévision
- 7) Le test de Ljung-Box pour l'autocorrélation des erreurs de mesure.

### III. Résultat

#### 1. SG

Le tableau 4 représente les premiers résultats obtenus sur un total de 1834 évaluations apportées aux 61 états de santé retenus par le processus de sélection orthogonal. Dans ce type d'approche, il n'existe pas d'interprétation possible des effets marginaux. Mais on s'attend à un signe négatif associé à une croissance des coefficients en valeur absolue à mesure que l'on évolue vers les niveaux élevés à l'intérieur d'un attribut (c.-à-d. les niveaux les plus sévères). À la première colonne, on retrouve le modèle des moindres carrés ordinaires simples (MCO) avec comme variables explicatives les 6 dimensions qui sont

présentes à travers leurs différents niveaux dans un ordre croissant. À l'intérieur d'une dimension, plus on se déplace vers les niveaux les plus élevés (de 1 vers 6), plus l'utilité associée à ces niveaux devrait diminuer. Ainsi le coefficient associé au niveau 5 doit être plus petit que celui associé au niveau 4 qui doit être plus petit que celui associé au niveau 3 et ainsi de suite. On s'attend à ce que cette décroissance de l'utilité se présente dans toutes les dimensions de toutes les spécifications que l'on aura à utiliser y compris celles pour la méthode du DCE. Rappelons-nous que tous ces coefficients sont attendus avec un signe négatif, car ils mesurent des impacts négatifs sur la santé parfaite. Ce qui pourrait se traduire par une croissance des coefficients en valeur absolue à mesure que l'on évolue vers les niveaux sévères à l'intérieur d'un attribut. Si cette condition n'est pas respectée, on est dans ce cas en présence d'une **incohérence**. À noter que lorsqu'un coefficient associé à un paramètre n'est pas significatif, son utilité est considérée comme étant nulle. Cette valeur nulle doit aussi être considérée dans ce processus de décroissance à l'intérieur de chaque attribut pour déceler d'éventuelles incohérences.

Sur le modèle qui représente les moindres carrés ordinaires (MCO), nous avons 2 incohérences et 10 paramètres significatifs sur un total de 25 qui sont estimés. Ces incohérences se situent à la dimension "RL" au passage du niveau 2 vers le niveau 3 et "SF" au passage du niveau 3 vers le niveau 4. Certains paramètres n'ayant pas rempli les conditions de départ ne sont pas incohérents puisque le niveau qui suit est associé à un coefficient non significatif (SF2 et PAIN2 colonne 1). Les niveaux 1 des attributs sont volontairement omis pour éviter des problèmes de colinéarité parfaite. Environ 60% des coefficients de cette régression ne sont pas significatifs et tous ces paramètres sont considérés comme étant nuls. Ce qui explique le nombre réduit des incohérences. Mais le modèle MCO présente des limites et cela se reflète à travers certains critères de choix (AIC et BIC). Le test de Hausman est en faveur d'une spécification à effet fixe que l'on retrouve en deuxième et troisième colonne. La différence réside sur le fait qu'à la 3<sup>ème</sup> colonne, la constante est contrainte à l'unité. Le nombre de paramètres significatifs augmente d'une unité à la 3<sup>ème</sup> colonne (effet fixe contraint) et de 2 unités à la 2<sup>ème</sup> colonne (effet fixe) par rapport au modèle MCO. Notons que le nombre d'incohérences reste exactement le même, c'est-à-dire 2 incohérences qui demeurent aux mêmes niveaux des mêmes attributs "RL"

(passage du niveau 2 au niveau 3), "*SF*" (passage du niveau 3 au niveau 4). Par contre une autre amélioration par rapport au modèle MCO se situe sur l'erreur absolue moyenne (MAE) qui passe de 32% à 22% dans le modèle contraint. On note aussi une décroissance des coefficients plus poussée au niveau des modèles à effet fixe notamment dans les attributs *SF*, *PAIN*, *MH* et *VIT*.

Malgré les divergences dans certains coefficients des attributs qui rapportent des signes contraires à nos attentes, les modèles d'estimations obtenus avec notre base de données illustrent un meilleur pouvoir explicatif global. Mis à part les modèles agrégés que nous verrons plus tard, tous ceux que nous avons déjà vus sont réalisés à l'échelle individuelle et parvenir à avoir un pouvoir explicatif aux alentours de 35% (MCO) et 42% (effet fixe) reste un résultat satisfaisant. Ce pouvoir explicatif tourne autour de 20% dans l'étude de Brazier et al. (2002). Et cette même statistique envoisine les 65% dans nos modèles agrégés qu'on retrouve dans les colonnes 4 à 7 du tableau 4. Ces derniers sont réalisés sur la base d'une agrégation à la moyenne (colonnes 4 et 6) ou à la médiane (colonnes 5 et 7) des différentes QALY associées aux états. Le nombre d'observations est égal au nombre d'états (61). Les modèles agrégés par la moyenne ne se portent pas mieux que leurs prédécesseurs MCO et effet fixe. Malgré un pouvoir explicatif élevé, ils accumulent la diminution des paramètres significatifs et augmentent le nombre d'incohérences. En dépit de la diminution de l'erreur moyenne absolue (MAE) qui devient très faible (moins de 10%), les modèles agrégés par la médiane présentent certaines irrégularités dans l'erreur de mesure notamment avec le test de JB qui signale une anormalité dans la distribution des erreurs de prévision. Par contre lorsque le modèle agrégé à la moyenne est contraint à l'unité (colonne 6) on retrouve des caractéristiques similaires à notre modèle de base (MCO). Il dispose même d'un avantage en matière de prévision dans la mesure où il se retrouve avec une erreur absolue moyenne de 7% contre 35% pour le modèle MCO. Ce modèle agrégé et contraint (colonne 6) enregistre une nette amélioration à ce niveau par contre pour ce qui est des incohérences, il en compte 3 c'est-à-dire plus que tous les autres modèles présents dans le tableau de régression SG sans termes d'interaction.

Tableau 4: régression SG sans termes d'interactions

	-1	-2	-3	-4	-5	-6	-7
	MCO	E,fixe	contrainte E,fixe	moyenne	médiane	contrainte moyenne	contrainte médiane
PF2	-0,0433 [0,0334]	-0,0264 [0,0342]	-0,0452 [0,0340]	-0,0226 [0,0574]	-0,0254 [0,0731]	-0,0876 [0,0617]	-0,0498 [0,0690]
PF3	0,0177 [0,0338]	0,0197 [0,0336]	-0,0048 [0,0331]	-0,00227 [0,0566]	-0,0813 [0,0720]	-0,0864 [0,0580]	-0,113* [0,0648]
PF4	-0,00618 [0,0375]	-0,0223 [0,0385]	-0,0447 [0,0383]	-0,0112 [0,0621]	-0,0298 [0,0790]	-0,0838 [0,0664]	-0,057 [0,0743]
PF5	-0,163*** [0,0375]	-0,154*** [0,0382]	-0,185*** [0,0374]	-0,130** [0,0613]	-0,220*** [0,0781]	-0,235*** [0,0606]	-0,259*** [0,0677]
RL2	-0,0862** [0,0404]	-0,0918** [0,0412]	-0,0923** [0,0414]	-0,0445 [0,0674]	-0,0271 [0,0858]	-0,0281 [0,0766]	-0,021 [0,0856]
RL3	0,0354 [0,0344]	0,0579* [0,0306]	0,044 [0,0305]	0,0113 [0,0499]	-0,039 [0,0635]	-0,0343 [0,0548]	-0,0561 [0,0612]
RL4	-0,00551 [0,0308]	-0,00541 [0,0299]	-0,0145 [0,0300]	0,00389 [0,0484]	0,0146 [0,0616]	-0,00464 [0,0551]	0,0114 [0,0616]
RL5	-0,101*** [0,0360]	-0,0619* [0,0325]	-0,0644** [0,0327]	-0,0624 [0,0529]	-0,103 [0,0674]	-0,0623 [0,0603]	-0,103 [0,0674]
SF2	0,0933*** [0,0338]	0,0631* [0,0343]	0,0570* [0,0344]	0,0493 [0,0574]	0,0167 [0,0730]	0,00626 [0,0637]	0,000592 [0,0712]
SF3	-0,0163 [0,0335]	-0,0256 [0,0342]	-0,0333 [0,0343]	-0,0559 [0,0570]	-0,161** [0,0726]	-0,111* [0,0624]	-0,181** [0,0697]
SF4	-0,0590* [0,0329]	-0,0867*** [0,0335]	-0,0976*** [0,0335]	-0,0491 [0,0551]	-0,0772 [0,0701]	-0,0956 [0,0608]	-0,0946 [0,0680]
SF5	-0,0585 [0,0367]	-0,0855** [0,0357]	-0,0871** [0,0359]	-0,0228 [0,0595]	-0,0254 [0,0757]	-0,0348 [0,0676]	-0,0299 [0,0756]

PAIN2	0,0721* [0,0426]	0,0826** [0,0412]	0,0515 [0,0406]	0,039 [0,0654]	0,0368 [0,0833]	-0,0717 [0,0648]	-0,00467 [0,0724]
PAIN3	-0,05 [0,0396]	-0,011 [0,0387]	-0,0382 [0,0382]	-0,0468 [0,0601]	-0,0683 [0,0764]	-0,141** [0,0609]	-0,103 [0,0680]
PAIN4	0,0396 [0,0420]	0,0296 [0,0394]	-0,00115 [0,0388]	0,0153 [0,0623]	-0,0136 [0,0793]	-0,105* [0,0586]	-0,0584 [0,0655]
PAIN5	-0,0186 [0,0516]	-0,0466 [0,0491]	-0,0743 [0,0488]	-0,0437 [0,0795]	-0,0793 [0,101]	-0,160* [0,0819]	-0,123 [0,0916]
PAIN6	-0,178*** [0,0463]	-0,207*** [0,0441]	-0,226*** [0,0440]	-0,120* [0,0703]	-0,165* [0,0894]	-0,198** [0,0757]	-0,194** [0,0846]
MH2	0,0301 [0,0340]	0,0389 [0,0327]	0,0325 [0,0329]	-0,0331 [0,0539]	-0,0819 [0,0686]	-0,0858 [0,0588]	-0,102 [0,0657]
MH3	-0,0265 [0,0338]	-0,0297 [0,0359]	-0,0476 [0,0358]	-0,0373 [0,0604]	-0,106 [0,0769]	-0,142** [0,0594]	-0,146** [0,0663]
MH4	0,0462 [0,0317]	0,0188 [0,0327]	0,011 [0,0328]	0,00953 [0,0535]	-0,0508 [0,0681]	-0,0449 [0,0582]	-0,0712 [0,0651]
MH5	-0,252*** [0,0460]	-0,234*** [0,0431]	-0,238*** [0,0433]	-0,151** [0,0728]	-0,145 [0,0927]	-0,164* [0,0828]	-0,15 [0,0926]
VIT2	0,0033 [0,0325]	0,0089 [0,0317]	0,00195 [0,0318]	0,00244 [0,0514]	0,0161 [0,0654]	-0,0149 [0,0583]	0,00964 [0,0651]
VIT3	-0,0388 [0,0329]	-0,0506 [0,0330]	-0,0725** [0,0326]	-0,0304 [0,0535]	-0,0201 [0,0681]	-0,105* [0,0557]	-0,0479 [0,0622]
VIT4	-0,0651** [0,0330]	-0,0654** [0,0321]	-0,0757** [0,0321]	-0,0641 [0,0525]	-0,0671 [0,0668]	-0,109* [0,0579]	-0,0838 [0,0647]
VIT5	-0,148*** [0,0354]	-0,132*** [0,0333]	-0,148*** [0,0332]	-0,124** [0,0544]	-0,170** [0,0692]	-0,196*** [0,0571]	-0,197*** [0,0638]



_cons	0,675*** [0,0574]	0,674*** [0,0546]	1 [,]	0,698*** [0,0881]	0,887*** [0,112]	1 [,]	1 [,]
N	1834	1834	1834	61	61	61	61
PS	10	12	11	4	4	11	6
INC.,	2	1	2	0	1	3	2
MAE	0,32	0,31	0,22	0,06	0,09	0,074	0,09
AIC	1617,6	739,6	1500,1	-72,5	-43,1	-56,9	-43,3
BIC	1761	883	3689,3	-17,63	11,8	-4,1	9,44
JB <sub>PRD</sub>	0	0		0	0,81	0,008	0,93
r <sup>2</sup>	0,35	0,426		0,601	0,655		

Standard errors in brackets

\* $p < 0,10$ , \*\* $p < 0,05$ , \*\*\* $p < 0,01$

Les nombres en couleur bleue illustre des incohérences

**INC.**: nombre d'incohérences, **JB<sub>PRD</sub>**: test de Jarque et Bera sur les erreurs prédites,

**MAE**: erreur absolue moyenne, **MH**: santé mentale, **N**: nombre d'observations, **PAIN**:

douleur, **PF**: fonctionnement physique, **PS**: nombre de paramètres significatifs, **RL**:

limitation des activités, **SF**: fonctionnement social, **VIT**: vitalité.

Le tableau 5 de régression avec l'inclusion des deux variables MOST et LEAST représente les mêmes caractéristiques de régressions que le tableau 4 à l'exception de l'ajout des termes d'interactions MOST et LEAST. Notons que toutes les régressions présentes dans le tableau 4 ont été associées aux variables MOST et LEAST, mais seules les meilleures (selon les critères de choix) ont été reportées. Il s'agit du modèle à effet fixe et du modèle agrégé à la moyenne. Cet ajout de nouvelles variables augmente le nombre de paramètres qui sont significatifs de 12 à 13, mais aussi le nombre d'incohérences qui passe de 1 à 4 dans la spécification à effet fixe. On parvient à corriger le passage du niveau 4 à 5 de la dimension SF qui fut un problème dans presque toutes les régressions sans interaction. Par contre, le passage du niveau 2 au niveau 3 devient un peu problématique dans 3 des 6 dimensions avec l'ajout des termes d'interaction. Mis à part les dimensions SF et VIT qui ne comptent que 2 paramètres significatifs (niveaux 4 et 5), tous les attributs renferment au moins une incohérence. L'erreur absolue moyenne et le pouvoir explicatif du modèle augmentent de 1%. On remarque en somme que l'ajout des termes d'interaction appauvrit les modèles en matière de nombres de paramètres significatifs et d'incohérences. Le même

impact est enregistré avec le modèle agrégé qui perd un peu de sa qualité (6 paramètres significatifs et 2 incohérences). Notons aussi qu'une seule variable (LEAST dans la régression à effet fixe) est significative parmi les variables qui ont été ajoutées dans les deux régressions. Le nombre de paramètres significatifs et le pouvoir explicatif du modèle sont en faveur du modèle à effet fixe avec les termes d'interactions. Par contre, le nombre d'incohérences et l'erreur absolue moyenne sont en faveur du modèle à effet fixe dépourvu des termes d'interaction. Si l'on ajoute le fait que la variable MOST n'est significative dans aucune des spécifications, on choisit le modèle à effet fixe sans termes d'interactions à des fins de comparaison avec la méthode du DCE.

Tableau 5. Régression SG avec l'ajout des variables d'interactions MOST et LEAST

	-1 E, fixe	-2 Contrainte Moyenne	-3 E, fixe	-4 Contrainte moyenne
PF2	-0,0264 [0,0342]	-0,0876 [0,0617]	<b>0,0104</b> <b>[0,0352]</b>	-0,0726 [0,0625]
PF3	0,0197 [0,0336]	-0,0864 [0,0580]	<b>0,0708**</b> <b>[0,0348]</b>	-0,0642 [0,0591]
PF4	-0,0223 [0,0385]	-0,0838 [0,0664]	0,00388 [0,0385]	-0,0599 [0,0672]
PF5	-0,154*** [0,0382]	-0,235*** [0,0606]	-0,0805** [0,0407]	-0,204*** [0,0634]
RL2	<b>-0,0918**</b> <b>[0,0412]</b>	-0,0281 [0,0766]	<b>-0,107**</b> <b>[0,0429]</b>	-0,0152 [0,0786]
RL3	<b>0,0579*</b> <b>[0,0306]</b>	-0,0343 [0,0548]	<b>0,0808**</b> <b>[0,0322]</b>	-0,0164 [0,0576]
RL4	-0,00541 [0,0299]	-0,00464 [0,0551]	-0,0126 [0,0341]	0,0167 [0,0626]
RL5	-0,0619* [0,0325]	-0,0623 [0,0603]	-0,0221 [0,0356]	-0,061 [0,0645]

SF2	0,0631* [0,0343]	0,00626 [0,0637]	0,0282 [0,0347]	0,0338 [0,0652]
SF3	-0,0256 [0,0342]	-0,111* [0,0624]	-0,0311 [0,0342]	-0,0914 [0,0635]
SF4	-0,0867*** [0,0335]	-0,0956 [0,0608]	-0,0599* [0,0346]	-0,0767 [0,0626]
SF5	-0,0855** [0,0357]	-0,0348 [0,0676]	-0,0652* [0,0360]	-0,0293 [0,0678]
PAIN2	0,0826** [0,0412]	-0,0717 [0,0648]	0,123*** [0,0423]	-0,0303 [0,0699]
PAIN3	-0,011 [0,0387]	-0,141** [0,0609]	0,0566 [0,0414]	-0,114* [0,0641]
PAIN4	0,0296 [0,0394]	-0,105* [0,0586]	0,0850** [0,0418]	-0,0612 [0,0657]
PAIN5	-0,0466 [0,0491]	-0,160* [0,0819]	-0,0156 [0,0490]	-0,119 [0,0846]
PAIN6	-0,207*** [0,0441]	-0,198** [0,0757]	-0,107** [0,0482]	-0,185** [0,0770]
MH2	0,0389 [0,0327]	-0,0858 [0,0588]	-0,00131 [0,0334]	-0,0556 [0,0611]
MH3	-0,0297 [0,0359]	-0,142** [0,0594]	0,00164 [0,0401]	-0,123* [0,0668]
MH4	0,0188 [0,0327]	-0,0449 [0,0582]	0,0466 [0,0334]	-0,0294 [0,0597]
MH5	-0,234*** [0,0431]	-0,164* [0,0828]	-0,214*** [0,0431]	-0,151* [0,0826]

VIT2	0,0089 [0,0317]	-0,0149 [0,0583]	-0,00281 [0,0320]	0,0013 [0,0588]
VIT3	-0,0506 [0,0330]	-0,105* [0,0557]	0,0102 [0,0348]	-0,0882 [0,0563]
VIT4	-0,0654** [0,0321]	-0,109* [0,0579]	-0,0709** [0,0334]	-0,0797 [0,0611]
VIT5	-0,132*** [0,0333]	-0,196*** [0,0571]	-0,0720** [0,0355]	-0,176*** [0,0586]
MOST			0,0203 [0,0506]	-0,0522 [0,0934]
LEAST			0,230*** [0,0461]	-0,0673 [0,0618]
_cons	0,674*** [0,0546]	1 [.]	0,314*** [0,0885]	1 [.]
N	1834	61	1834	61
PS	12	11	13	6
INC.,	1	4	3	2
MAE	0,31	0,074	0,32	0,068
AIC	739,6	-56,9	709,3	-57,78
BIC	883	-4,1	863,7	-0,79
JB <sub>PRD</sub>	0	0,008	0	0
r <sup>2</sup>	0,426		0,436	

Standard errors in brackets

\* $p < 0,10$ , \*\* $p < 0,05$ , \*\*\* $p < 0,01$

Les nombres en couleur bleue illustre des incohérences.

## 2. DCE

La méthode du DCE quant à elle aboutit à des résultats plus satisfaisants avec pas moins de 22 paramètres significatifs dans toutes les régressions du tableau 6. Le nombre d'incohérences ne dépasse pas 3 avec une erreur absolue moyenne de 3.1 qui n'est pas comparable avec le SG vu la procédure de chaque méthode. Le passage du niveau 2 au

niveau 3 dans l'ensemble des attributs renferme 4 des 5 incohérences qui sont présentes dans les régressions sans termes d'interactions (colonnes 1 et 2). Cette méthode accorde un poids important aux différents niveaux des attributs. La décroissance est beaucoup plus prononcée. Si on va plus loin on peut même avancer que les individus sont très sensibles aux attributs PAIN et MH dont les valeurs des niveaux les plus sévères fluctuent autour de  $[-1.142 ; -1.099]$  et  $[-1.184 ; -0,976]$ <sup>3</sup> respectivement. Les deux modèles que sont le LOGIT conditionnel (colonne 1) et le LOGIT à paramètres aléatoires (colonne 2) se valent même si les tests AIC et BIC sont légèrement en faveur du premier. Ils se retrouvent tous les deux avec le même nombre de paramètres significatifs (22), 3 incohérences pour le LOGIT conditionnel contre 2 pour le Logit aléatoire et une erreur absolue moyenne de 2.32 et 2.57 respectivement.

Les deux dernières colonnes représentent des régressions de modèles LOGIT conditionnels et aléatoires ajoutés de termes d'interactions. Cette fois, ils sont tous significatifs et améliorent nos régressions contrairement à la méthode du SG. Ils gardent le même nombre d'incohérences que dans nos deux régressions de bases et améliorent le nombre de paramètres significatifs qui passe à 2 unités additionnelles dans le LOGIT conditionnel et à 3 unités dans celui aléatoire. Le passage du niveau 2 au niveau 3 ne renferme pratiquement que des incohérences pour les attributs PF, RL et PAIN. Pour la dimension PF, le passage du niveau 2 au niveau 3 n'est pas significatif dans la régression du LOGIT conditionnel, par contre il est significatif et incohérent dans le Logit aléatoire. Ce qui est contraire à ce que l'on observe à la dimension PAIN qui n'est pas significative avec le Logit aléatoire et significatif et incohérent avec LOGIT conditionnel toujours avec le même niveau. La dimension RL qui regroupe uniquement des paramètres significatifs enregistre une incohérence au passage du niveau 2 au niveau 3 dans toutes les régressions. La transition du niveau 2 vers le niveau 3 reste visiblement problématique et cela est généralement dû au niveau 2 des attributs que les individus semblent avoir du mal à évaluer.

La présence des termes d'interaction crée une décroissance plus poussée par rapport aux modèles qui en sont dépourvus avec un maximum de **-1.184** que l'on enregistre au niveau

---

<sup>3</sup> Ces intervalles sont obtenus en considérant le minimum et le maximum d'un attribut à travers les différentes régressions. Il en sera ainsi pour tous les intervalles de fluctuations.

6 de la dimension PAIN. En d'autres termes, la méthode du DCE associé au Logit aléatoire nous dit que le fait de trouver le niveau 6 de la dimension PAIN dans un état de santé est suffisant pour le qualifier de **pire que la mort**<sup>4</sup>. Le même constat peut être fait avec l'attribut MH qui tourne aux alentours de **-1.142**. Notons aussi que ces modèles sont accompagnés d'erreurs absolues moyennes assez grandes comparées à celles que l'on retrouve dans la méthode du SG qui enregistre un maximum de 0,32.

La confrontation entre le modèle à effet fixe (SG) et du LOGIT conditionnel (DCE) sans termes d'interactions regroupe plusieurs distinctions. Tout d'abord, il faut savoir que cette confrontation est le résultat de tout ce qui a été fait dans les deux méthodes. Il ne s'agit pas de comparer uniquement deux régressions, mais plutôt tout le processus qui a été fait pour aboutir à ces régressions. Le nombre de paramètres significatifs que l'on retrouve dans le LOGIT conditionnel avec le DCE représente presque le double de ce que l'on voit sur le modèle à effet fixe avec SG. Le contraire est observé dans les incohérences où ce dernier arrive avec seulement une incohérence contre 3 pour le LOGIT. Mais une explication peut être avancée pour comprendre le nombre d'incohérences notamment assez faible dans le modèle à effet fixe (SG). Les paramètres significatifs de ce modèle se concentrent en majorité sur les niveaux les plus sévères (4 ou 5). Dans les deux modèles, les niveaux 5 sont toujours négatifs et les répondants semblent montrer plus d'aptitude à faire la distinction entre les passages des niveaux 4 à 5 que ceux allant de 2 à 3. Ce faisant il sera difficile de trouver une incohérence si la majorité des paramètres associés aux autres niveaux (2 et 3) ne sont pas significatifs.

Une incohérence est définie lorsque le passage d'un niveau vers un autre plus élevé ne se reflète pas sur les coefficients des paramètres en manière de décroissance, mais jusqu'à présent, aucune mention n'a été faite lorsque les deux niveaux qui se suivent à l'intérieur d'un attribut sont associés à des coefficients non significatifs. Dans un tel cas, ces paramètres, bien qu'ayant une valeur, sont considérés comme nuls et devraient aussi constituer une incohérence puisque le passage d'un niveau à l'autre n'est plus décroissant,

---

<sup>4</sup> On fait l'hypothèse que la valeur d'un état est donnée par l'équation  $1 - \sum A_{IJ}$  avec  $-\sum A_{IJ}$  la somme des coefficients associés aux différents niveaux d'un état. Si  $-\sum A_{IJ}$  est plus petit que -1, l'état en question est pire que la mort. 1 étant le QALY de la santé parfaite, tous les états sont déduits de ce dernier.

mais constant. Ainsi le modèle à effet fixe de la méthode SG semble avoir plus de problèmes en matière d'incohérences que ce qui a été reporté dans les tableaux 4 et 5.

On a aussi une différence très marquée dans la décroissance des paramètres qui est largement plus importante dans le LOGIT. Lorsqu'il n'y a pas d'incohérences aux premiers niveaux, les paramètres sont assez proches en termes d'utilité en général, mais la tournure survient assez rapidement à partir des niveaux 3. Et cette différence d'utilité est encore plus marquée dans les niveaux permettant de définir les pires états de santé. Mise à part l'erreur absolue moyenne (MAE), tous les critères de choix que l'on a eu à considérer sont en faveur du LOGIT conditionnel qui utilise les données issues de la méthode DCE.

Tableau 6. Régression DCE (LOGIT)

	-1	-2	-3	-4
	conditionnel	Multinomial	conditionnel	Multinomial
main				
PF2	-0,0463 [0,0553]	<b>-0,240***</b> <b>[0,0560]</b>	-0,0259 [0,0559]	<b>-0,222***</b> <b>[0,0568]</b>
PF3	-0,0418 [0,0555]	<b>-0,159***</b> <b>[0,0571]</b>	-0,0567 [0,0564]	<b>-0,164***</b> <b>[0,0563]</b>
PF4	-0,361*** [0,0548]	-0,710*** [0,0624]	-0,391*** [0,0553]	-0,747*** [0,0625]
PF5	-0,596*** [0,0572]	-0,920*** [0,0674]	-0,612*** [0,0574]	-0,943*** [0,0672]
RL2	<b>-0,531***</b> <b>[0,0554]</b>	<b>-0,565***</b> <b>[0,0590]</b>	<b>-0,549***</b> <b>[0,0558]</b>	<b>-0,600***</b> <b>[0,0585]</b>
RL3	<b>-0,194***</b> <b>[0,0558]</b>	<b>-0,264***</b> <b>[0,0593]</b>	<b>-0,224***</b> <b>[0,0564]</b>	<b>-0,284***</b> <b>[0,0589]</b>
RL4	-0,395*** [0,0551]	-0,442*** [0,0606]	-0,434*** [0,0564]	-0,471*** [0,0611]

RL5	-0,582*** [0,0565]	-0,676*** [0,0650]	-0,647*** [0,0589]	-0,761*** [0,0667]
SF2	-0,0271 [0,0511]	-0,0967* [0,0529]	-0,0328 [0,0511]	-0,134** [0,0530]
SF3	<b>-0,247***</b> <b>[0,0513]</b>	-0,206*** [0,0541]	<b>-0,261***</b> <b>[0,0515]</b>	-0,204*** [0,0538]
SF4	<b>-0,130**</b> <b>[0,0520]</b>	-0,299*** [0,0549]	<b>-0,175***</b> <b>[0,0533]</b>	-0,340*** [0,0576]
SF5	-0,407*** [0,0528]	-0,599*** [0,0593]	-0,448*** [0,0538]	-0,681*** [0,0626]
PAIN2	<b>-0,230***</b> <b>[0,0530]</b>	0,00488 [0,0553]	<b>-0,243***</b> <b>[0,0530]</b>	-0,0621 [0,0573]
PAIN3	<b>0,124**</b> <b>[0,0552]</b>	0,0565 [0,0612]	<b>0,0805</b> <b>[0,0562]</b>	-0,097 [0,0698]
PAIN4	-0,325*** [0,0547]	-0,273*** [0,0585]	-0,367*** [0,0562]	-0,375*** [0,0658]
PAIN5	-0,709*** [0,0628]	-0,820*** [0,0726]	-0,751*** [0,0639]	-0,923*** [0,0780]
PAIN6	-0,976*** [0,0615]	-1,101*** [0,0883]	-1,018*** [0,0624]	-1,184*** [0,0922]
MH2	-0,137*** [0,0488]	-0,199*** [0,0560]	-0,122** [0,0492]	-0,166*** [0,0614]
MH3	-0,456*** [0,0538]	-0,266*** [0,0573]	-0,426*** [0,0579]	-0,210*** [0,0657]
MH4	-0,675*** [0,0523]	-0,523*** [0,0564]	-0,726*** [0,0544]	-0,578*** [0,0594]



MH5	-1,128*** [0,0533]	-1,107*** [0,0683]	-1,142*** [0,0536]	-1,099*** [0,0666]
VIT2	-0,112** [0,0499]	-0,0197 [0,0529]	-0,105** [0,0500]	-0,0419 [0,0558]
VIT3	-0,299*** [0,0519]	-0,243*** [0,0564]	-0,310*** [0,0529]	-0,267*** [0,0626]
VIT4	-0,540*** [0,0504]	-0,563*** [0,0544]	-0,568*** [0,0511]	-0,625*** [0,0573]
VIT5	-0,666*** [0,0528]	-0,673*** [0,0625]	-0,707*** [0,0540]	-0,782*** [0,0681]
MOST			0,220*** [0,0840]	0,309*** [0,108]
LEAST			-0,302*** [0,0935]	-0,664*** [0,153]
_cons	0,874*** [0,157]	0,985*** [0,181]	0,904*** [0,163]	1,012*** [0,206]
N	17320	17320	17320	17320
PS	22	22	24	25
INC.,	3	2	3	2
MAE	2,32	2,57	2,51	3,1
AIC	19031,5	10153,7	19019,4	10135,9
BIC	19225,5	10440,8	19228,9	10438,5
ML	-9490,7	-5059,8	-9482,7	-5028,9
JB <sub>PRD</sub>				
Pseudo r <sup>2</sup>	0,0946		0,0954	

Standard errors in brackets

\* $p < 0,10$ , \*\* $p < 0,05$ , \*\*\* $p < 0,01$

Les nombres en bleu illustrent des incohérences

## IV. Discussion

Les résultats de cette étude vont permettre la comparaison entre le SG et le DCE. À son issue, on détermine la meilleure méthode à utiliser pour le calcul du QALY des états de santé et une analyse coût-utilité adéquate. La détermination du QALY d'un état de santé dépend généralement de 3 facteurs. Nous examinons dans les parties suivantes ces 3 éléments dans chaque méthode pour avoir la meilleure approche pour le QALY :

### *1- L'instrument*

Les deux méthodes utilisent le SF6Dv2. Aucune distinction ne peut être faite à ce niveau entre le SG et DCE.

### *2- La méthode d'évaluation*

En comparant les deux méthodes au niveau de leur structure et de leur administration durant l'enquête, le DCE a légèrement l'avantage de la simplicité dans la procédure. Et qui parle de simplicité fait allusion à la compréhension. Il a toujours été question de faire des choix pour déterminer le QALY des états de santé, mais la procédure des deux méthodes n'est pas la même.

La littérature présente le DCE comme étant la méthode la plus simple comparée au SG, mais dans cette étude, du côté du SG comme le DCE les individus jugent en général l'épreuve difficile (l'avantage du DCE est faible à ce niveau). La difficulté dans le SG semble se situer sur la manière d'évaluer un état avec le pingpong. Pour le DCE, c'est le choix entre deux états en absence de dominance tout en considérant autant d'attributs qui est contraignant. Toutes les deux procédures ont été réalisées en ligne, mais l'impact de l'évaluation sera différent selon la méthode. Puisque la difficulté est légèrement au désavantage du SG, réaliser des enquêtes en ligne aura plus d'impacts négatifs sur cette méthode qui demande davantage d'explications. Nous observons que tout au long de l'enquête, beaucoup d'incompréhensions ont été commises comme le fait de choisir un traitement alors que les chances de succès sont nulles ou même le refuser alors que les chances de succès sont de 100%. On a aussi relevé des commentaires de certains individus disant qu'ils avaient de la difficulté à comprendre certaines tâches au début, mais que c'est

avec la répétition qu'ils sont parvenus à mieux comprendre le SG. Pour le DCE, l'application d'une enquête en ligne peut renfermer moins de problèmes. En effet, faire un choix entre deux alternatives définies à l'aide de 6 critères (SF6Dv2) et la durée de vie peut-être assez délicat surtout que l'on a pris le soin d'exclure toute sorte de dominance entre les attributs. Est-ce possible de considérer autant de critères (6 dimensions plus la durée) dans un choix ? La réponse peut être affirmative, mais la résultante est une difficulté assez élevée.

Dans les tableaux de régression SG, plus de la moitié des coefficients de paramètres ne sont pas significatifs. La totalité des coefficients non significatifs se situe dans les niveaux 2 et 3 des attributs. Dans la littérature, les auteurs ont souligné les innombrables interrogations et problèmes que renferme le SG. Parmi eux, on peut citer le problème de la distribution non linéaire de la probabilité (Van Houtven 2011) le risque de perte (Bleichrodt 1999), la neutralité vis-à-vis du risque (Bleichrodt 2002). Rappelons que dans les caractéristiques des données on a trouvé que tous les états ont été évalués au moins une fois comme ayant un QALY de 1 (santé parfaite). Un individu disposant d'un état de santé imparfait et qui refuse un traitement pouvant le mener à la santé parfaite à 100% semble avoir une aversion au risque ou a du mal à comprendre le but de l'exercice. Des auteurs comme (Van Houtven 2007) et (Bleichrodt 1999) associent ce fait à une défaillance de la théorie et au comportement de l'aversion pour le risque. Brazier et al. (2002) rapporte que les répondants surestiment la valeur des états faibles. La considération de tous ces faits signifie que nos résultats (SG) sur la significativité peuvent être en partie la résultante des limites soulevées dans la théorie. Une autre approche de nos résultats du SG est de considérer le risque comme étant la cause principale de l'absence de significativité dans les paramètres. La majeure partie des paramètres significatifs sont situés dans les niveaux d'attributs les plus faibles. Ainsi notre variable expliquée qui représente le QALY d'un état (probabilité) serait mieux expliquée par les niveaux faibles (4, 5, et 6) des attributs. Il y aurait une corrélation négative entre la qualité d'un état de santé et l'utilité des niveaux faibles (4, 5, et 6) des attributs. Cette corrélation devrait être négative avec tous les niveaux des attributs si la neutralité au risque était respectée. Une des explications que l'on pourrait avancer est que les individus préfèrent ne pas agir lorsqu'ils ont en leur possession un état

intermédiaire qui est satisfaisant, mais reste assez sensible à un traitement au moment où la santé se dégrade (état faible).

La difficulté peut être évitée en optant pour une enquête menée par des individus expérimentés prêts à interagir avec les répondants au besoin. Il ne sera pas de même pour l'aversion au risque qui est présente, *mais notons que cette aversion pour le risque ne semble pas affecter les niveaux les plus élevés.*

À ce niveau le SG est légèrement avantageé puisque la difficulté peut être réduite en considérant des enquêtes menées par des individus expérimentés et *la neutralité vis-à-vis au risque semble être respectée pour les niveaux faibles.*

### 3- Le modèle

L'ensemble des spécifications utilisées dans cette étude renferme des distinctions à l'application du calcul de QALY. Du côté du SG on note une absence de significativité (12/25), une erreur absolue moyenne faible (31%) et un pouvoir explicatif plutôt élevé (42,5) et similaire à celui du modèle de York MVH à l'application du EQ-5D Dolan et al (1997). Pour ce qui est des modèles individuels, le pouvoir explicatif reste bien au-dessus de ceux de l'étude de Brazier et al (2002), et ce malgré le nombre d'incohérences et l'absence de significativité au niveau des coefficients. La majorité des incohérences dans la méthode du SG se concentrent sur 2 attributs : RL et SF. On constate que ces deux dimensions sont moins précises dans l'information qu'elles apportent à la définition d'un état de santé par rapport aux autres dimensions. Exemple. « **Jamais accomplir moins de choses** » est un exemple de déclaration qu'on peut trouver dans un état de santé pour la dimension RL *limitation du rôle*. « **Activités sociales jamais limitées** » est un autre exemple de la dimension SF *fonctionnement social* qui se focalise sur les activités sociales qui peuvent être nombreuses et variées. Ces attributs sont moins précis et en l'occurrence plus difficiles à comprendre pour les répondants.

Beaucoup d'auteurs ont conclu que les individus ont tendance à sous-estimer la valeur des pires états de santé par exemple Brazier et al (2002). Mais dans nos résultats du SG, la majeure partie des niveaux qui sont significatifs se localisent dans les niveaux faibles. Les niveaux 5 et 6 sont significatifs dans toutes les dimensions sur l'ensemble des régressions.

Si l'erreur absolue moyenne est un peu élevée par rapport à la valeur enregistrée dans Brazier et al. (2002), cela devrait être dû à l'absence de significativité qui caractérise la majorité des paramètres que l'on retrouve dans le tableau de régression SG. Malgré cette absence de significativité dans les paramètres, le pouvoir explicatif des modèles de régression du SG dans les tableaux 4 et 5 reste élevé et l'erreur absolue moyenne faible par rapport au DCE.

Cependant, attirons notre attention sur un fait : Le MAE enregistré dans le modèle SG est très faible par rapport à celui du DCE, mais reste plus grand que celui de Brazier et al. (2002). Par contre le pouvoir explicatif de ce même modèle SG est 2 fois plus élevé que celui de Brazier et al. (2002) en dépit de l'absence de significativité enregistré dans notre étude. Cela pourrait appuyer la neutralité du risque par rapport aux niveaux faibles qui semblent être respectés dans nos données. *Autrement dit, l'absence de significativité dans les niveaux 2 et 3 des attributs serait bénéfique dans la mesure où la neutralité au risque n'est pas respectée pour ces niveaux.*

Pour le DCE, les modèles affichent de bons résultats en matière de paramètres significatifs et de décroissance à part quelques incohérences dans la dimension fonctionnement physique (PF), limitation des activités (RL) et fonctionnement social (SF) comme dans le SG, mais particulièrement le niveau 2 que les modèles ont du mal à évaluer. On pourrait avancer la même explication que celle de la méthode du SG sur l'ambiguïté des attributs SF et RL. Mais pour le moment, nous avons deux questions sans réponse. Pourquoi les individus ont-ils du mal à évaluer le niveau 2 de la dimension RL dans le DCE? Pourquoi les utilités sont-elles aussi grandes en valeur absolue dans l'ensemble des niveaux de tous les attributs ?

Pour avoir une explication, il faut remonter plus haut dans le questionnaire. À la fin de l'exercice du DCE, on demande aux individus le nombre de dimensions qu'ils ont considéré au moment du choix.

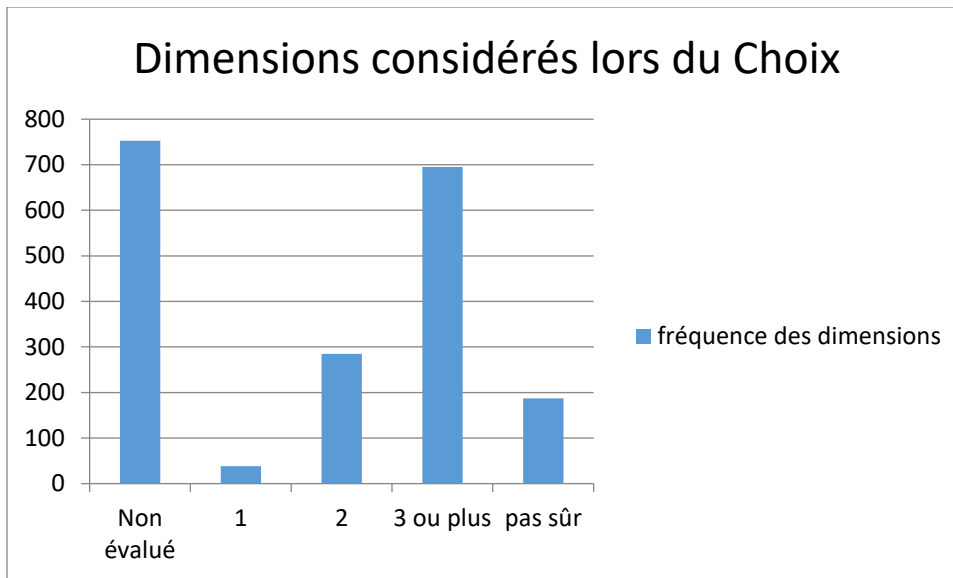


Figure 2 : Nombre d'attributs considérés au moment du choix dans le DCE

Comme on peut le constater, la majorité de ceux qui ont répondu dit avoir considéré 3 ou plus de dimensions.

Maintenant, si l'on se concentre sur les 3 dimensions auxquelles on accorde le plus d'importance, on se retrouve en première position avec la durée, la douleur (PAIN) et la santé mentale (MH). Si on fait une liste des états santé dont la dimension SF est représentée par le niveau 2, on constate que sur 72% de ces états (18 sur 25), ce niveau 2 est associé aux pires niveaux de la dimension de la douleur et la santé mentale. Ce qui fait que l'évaluation de ce niveau sera conditionnée par la présence des pires niveaux des attributs PAIN et MH.

Quasiment la majorité des coefficients obtenus dans la méthode de DCE sont significatifs avec le signe négatif attendu. Par contre, les modèles de régression dans cette méthode enregistrent une erreur absolue moyenne extrêmement grande due à la taille des coefficients, mais aussi à la nature de la variable expliquée (le choix entre alternatives transformées en variable dichotomique 0/1).

Les différents modèles que l'on a utilisés dans cette étude apportent des résultats différents selon les critères de choix considérés dans la détermination du QALY d'un état de santé.

Il nous a été parfois difficile de porter un choix entre les modèles selon le critère d'Akaike par exemple lorsqu'on ajoute les termes d'interactions. Mais finalement, l'importance que l'on accorde à une méthode dépend de sa capacité à prédire le QALY des états du SF6Dv2.

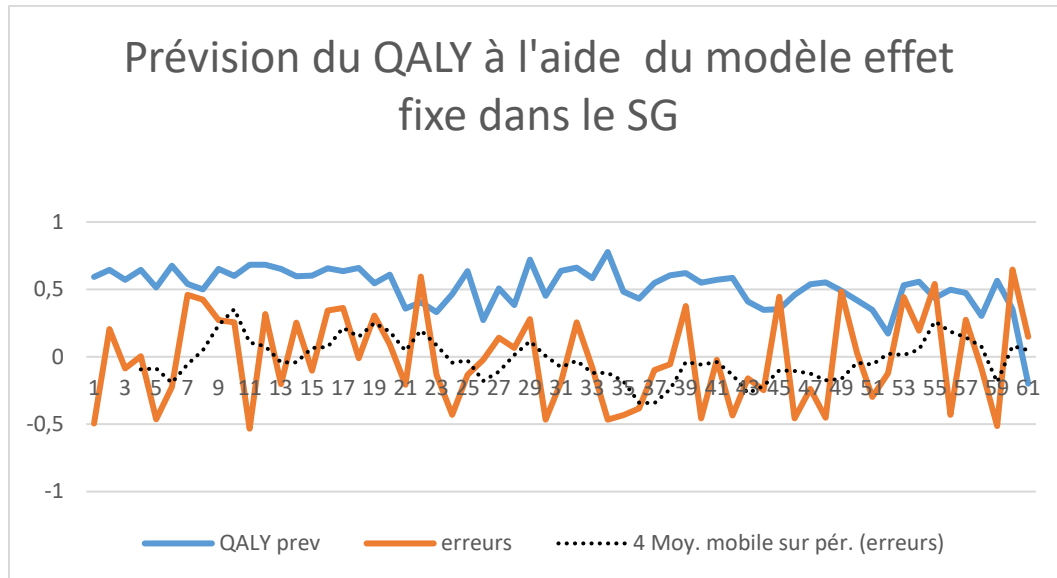


Figure 6 : Prédiction du QALY à l'aide du SG

Le QALY prédit est calculé en faisant d'abord une prévision linéaire à partir du modèle à effet fixe choisi pour le SG. Par la suite on calcule une moyenne de ces prévisions pour chaque état de santé. L'erreur représente la différence moyenne entre la valeur réelle et la valeur prédite. La moyenne mobile d'ordre 4 est choisie pour réduire les risques d'autocorrélation entre les 3 attributs que les individus rapportent avoir le plus considérés.

La figure 2 représente une prévision du QALY des 61 états de santé utilisés dans l'épreuve du SG. Le QALY se situe dans l'intervalle de 0,3 à 0,6 en moyenne. Ce résultat est conforme aux valeurs de QALY des états obtenus dans le tableau 2 lors de la description de ces mêmes 61 états. Trois problèmes majeurs peuvent être soulevés :

Le premier est la difficulté à prédire les faibles<sup>5</sup> états de santé. En effet, on note un décalage plus prononcé entre la courbe de prévision et celle de la tendance pour ce qui est de la prévision des états faibles.

<sup>5</sup> Un état de santé est qualifié de faible dans cette étude lorsqu'il est défini avec les niveaux élevés des attributs du SF6Dv2 notamment ceux que les individus rapportent avoir le plus considérés au moment du choix.

Le second est la grande marge de fluctuation des erreurs qui se situent dans l'intervalle -0,5 et 0,6. Ces erreurs suivent une marche aléatoire et peuvent atteindre une valeur très grande.

Le 3<sup>em</sup> aspect est le test de Ljung-Box d'autocorrélation des résidus qui révèle une corrélation significative des erreurs prédites.

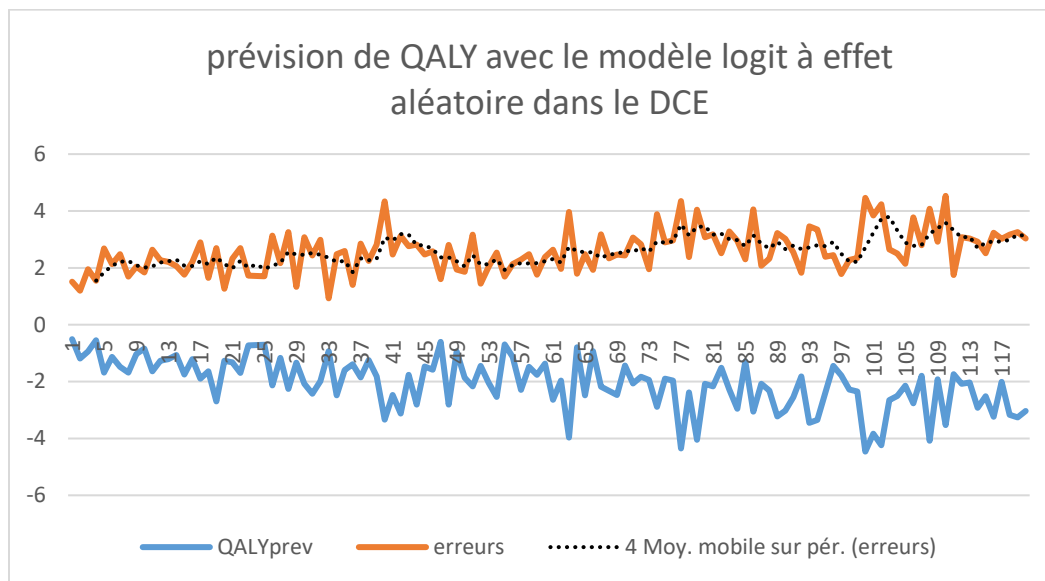


Figure 7 : prévision du QALY avec la méthode du DCE (Logit conditionnel)

Le même procédé utilisé dans le SG est utilisé dans le DCE pour obtenir la figure 3.

La figure 7 représente une prévision du QALY des 120 états de santé utilisés dans l'exercice du DCE. Les améliorations obtenues dans cette méthode en matière de nombre de paramètres significatifs en matière de nombre d'incohérences sont contrebalancées par la décroissance des paramètres notamment enregistrés à la dimension 2 de l'activité sociale et des dimensions 5 et 6 de tous les autres attributs. Ce modèle prédit le QALY des états de santé dans l'intervalle -1; -4 et l'erreur de mesure entre 2 et 4. En dépit de toutes les limites qu'il présente, le SG arrive à mieux prédire les états de santé. Les résultats obtenus par la méthode du DCE ne sont pas acceptables en matière de prévision. Mais cette difficulté de prévision est explicable à l'aide de l'importance de l'erreur absolue moyenne dans cette méthode qui est due à la nature de la variable expliquée dans le DCE. L'ensemble des états qui ont fait l'objet de cette étude ne peuvent pas avoir un QALY négatif. Cependant, un aspect important est à noter :



La méthode du DCE semble mieux décrire les erreurs de prévision comme une variable stationnaire (une variable qui fluctue autour de sa moyenne).

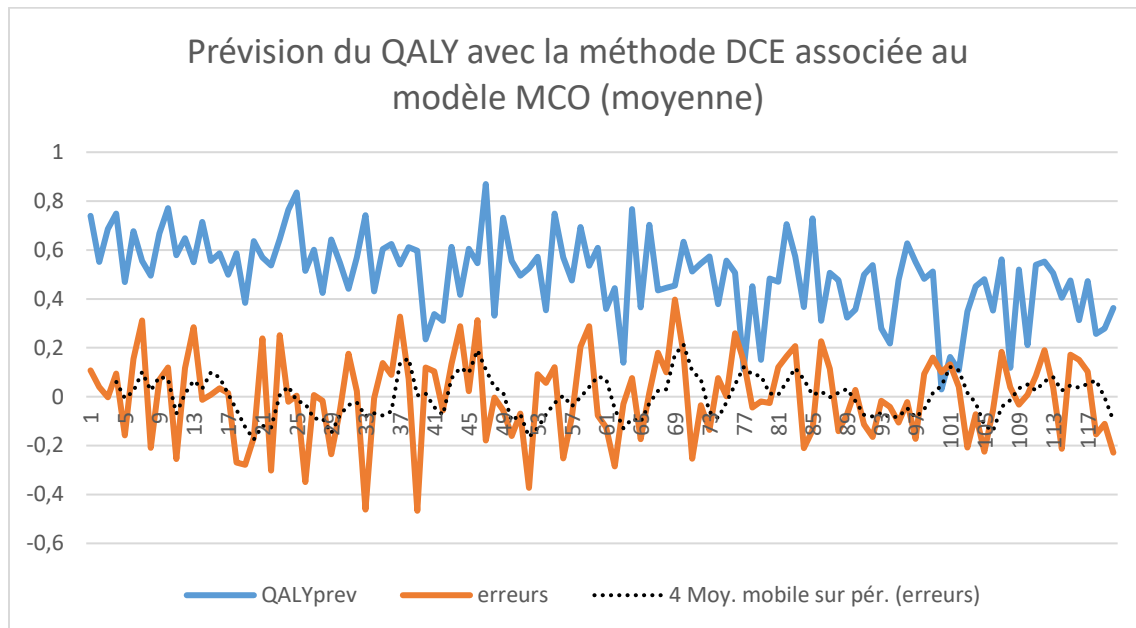


Figure 8 : prévision du QALY à l'aide du DCE associé au MCO

La figure 8 représente toujours la méthode du DCE. En effet, son avantage soulevé précédemment nous a poussés à reconsidérer les résultats obtenus dans cette méthode. Le principal problème que l'on trouve avec cette méthode est le poids des erreurs de mesure (MAE). Le DCE présente une difficulté moins importante que le SG dans la procédure et arrive à fournir des résultats satisfaisants dans la significativité des paramètres et du nombre d'incohérences. Puisque le modèle MCO ne présente pas de problèmes liés à l'erreur de mesure, on prend la décision de l'associer à la méthode du DCE. Cette procédure ne garantit pas que la variable explicative sera comprise entre 0 et 1 et les coefficients obtenus ne seront pas optimaux dans la mesure où le terme d'erreur n'est pas homoscédastique. Pour contourner ce problème, on transforme la variable expliquée binaire en une moyenne pour chaque état de santé avant de l'estimer par les moindres carrés ordinaires (voir annexe 4). Autrement dit, on fait un modèle agrégé à la moyenne à partir de la variable binaire. Les résultats en matière de prévision sont reportés à la figure 4. La prévision du QALY des états de santé varie de 0,7 et diminue régulièrement d'une manière plus lisse jusqu'à 0,2. Les résidus sont toujours stationnaires, mais cette fois autour de l'axe des abscisses (autour de la valeur de zéro) ce qui est très bien pour la prévision de

l'ensemble des états qui compose le SF6Dv2. Ainsi, on a trouvé un moyen d'associer à la méthode DCE une spécification qui est en mesure de réduire les résidus et les résultats sont meilleurs que le DCE associé au Logit conditionnel ainsi que le SG associé au modèle effet fixe en matière de prévision. On voit l'importance d'associer une bonne spécification à la méthode. Ce dernier peut avoir d'énormes avantages dans l'inclusion des préférences sur la détermination du QALY, mais encore faudrait-il trouver le bon modèle capable de les montrer.

Tous les critères que l'on a considérés étaient parfois en faveur du SG, parfois au côté du DCE, mais la prévision semble avoir donné l'avantage au DCE associé au modèle MCO pris en moyenne dans la prévision du QALY des états que renferme le SF6Dv2.

Cependant, le modèle agrégé qui est associé au DCE enregistre des caractéristiques qui se rapprochent beaucoup plus au modèle effet fixe dans le SG que du Logit dans le DCE. On ne retrouve pas plus de 10 paramètres significatifs avec une seule incohérence au passage du niveau 2 vers le niveau 3 de la dimension SF. Le pouvoir explicatif du modèle est de 48%. Il semble que le choix de l'individu est en fonction de la présence des niveaux permettant de définir les pires états de santé et peu importe la méthode.

Dans le SG, ce fait est associé à la neutralité au risque vis-à-vis de ces niveaux. De ce fait, seuls les états faibles ou jugés comme tels par le répondant seront en mesure d'être évalués avec neutralité vis-à-vis du risque. En effet, lorsque l'état évalué est jugé satisfaisant le concerné considèrera toujours ce risque de perte pouvant le mener au PITS. Par contre lorsque le répondant juge l'utilité de l'état évalué pas trop différent de celle du PITS, cette aversion pour la perte n'a plus lieu d'être. Ce qui entraîne que l'hypothèse de neutralité au risque n'existe plus. Par exemple lorsque l'individu évalue un état défini par les niveaux 545645. Il peut juger dans ce cas l'utilité de son état pas trop différent de celui du PITS entraînant ainsi la neutralité au risque envers les niveaux de l'état de santé précédent.

Pour le DCE, nous retiendrons que la méthode est fortement corrélée à la spécification qui lui est associée. Ainsi, la loi de probabilité du LOGIT conditionnel ainsi que celle du modèle LOGIT à paramètre aléatoire ne semblent pas appropriées au QALY.

À noter que ces conclusions sont tirées à partir des données que nous avons utilisées dans cette étude.

## Conclusion

La littérature laisse entendre que le QALY est un instrument de plus en plus utilisé. Les différentes méthodes utilisées pour son calcul ont comme principal but d'inclure les préférences des individus de manière cohérente et logique. Mission qui n'est pas aisée puisque dans les deux méthodes que nous avons utilisées dans cette étude, chacune d'elle présente des avantages comme des inconvénients. Le SG est la seule méthode qui se base dans un univers probabiliste, par contre l'hypothèse de neutralité vis-à-vis du risque n'est pas respectée selon la littérature. Il est jugé difficile et fait appel à un travail fastidieux pour sa mise en place. Le DCE qui est basé sur un simple classement des états de santé est jugé plus simple, mais par contre ne respecte pas l'hypothèse d'indépendance des alternatives non pertinentes. Les données récoltées dans cette étude provenant de la population du Québec rapportent que cette difficulté est présente dans chacune des méthodes. Un peu plus de la moitié des individus assignés à chaque méthode jugent l'exercice (SG et DCE) difficile. Raison pour laquelle dans chaque méthode, les données ont été triées et les réponses incohérentes supprimées. 13% de l'échantillon du SG ont été supprimés pour des raisons d'incompréhension. Au niveau du DCE 20% ont été retirés pour la même raison. Les individus exclus présentent plus de problèmes liés à la santé que ceux qui sont inclus dans les deux méthodes. Mais cela n'empêche que celles qui sont utilisées dans les modèles de régression présentent une grande diversité qui témoigne de la complexité des préférences dans le QALY.

Ces modèles que nous avons utilisés dans la tâche de détermination de QALY sont nombreux et divers. Chacune d'elle a tenté d'expliquer au mieux le QALY d'un état de santé à partir d'éléments observables (SF6Dv2). L'ensemble des spécifications utilisées, dont l'erreur absolue moyenne est aux alentours de 10%, ont un nombre de paramètres significatifs compris entre 10 et 14. Le seul modèle dont le nombre de paramètres significatifs dépasse de loin cet intervalle est le LOGIT (conditionnel ou multinomiale) utilisé dans le DCE. Dans le SG, pratiquement seuls les niveaux faibles sont significatifs

et sans aucune incohérence dans l'ensemble des attributs. L'aversion au risque ne semble pas affecter ces niveaux. En d'autres termes, le SG fournit des résultats biaisés pour ce qui est de l'évaluation des niveaux 2 et 3 (surtout le niveau 2). Ces niveaux sont non significatifs dans le QALY des états de santé du SF6Dv2 ou lorsqu'ils le sont, ils deviennent incohérents dans leur détermination. Pour ce qui est du DCE, la loi de probabilité du LOGIT conditionnel ou du LOGIT multinomial à coefficients aléatoires n'est pas adaptée aux données de cette étude. L'utilité des différents niveaux des attributs est très élevée en valeur absolue. La résultante est une erreur absolue moyenne très grande. Si l'on insiste autant sur le nombre de paramètres significatif, le nombre d'incohérences, le pouvoir explicatif d'un modèle et l'erreur absolue moyenne c'est parce le choix d'une méthode va dépendre de sa capacité de prévision.

À ce niveau, le DCE qui est associé au MCO après que la variable expliquée binaire est transformée en moyenne fournie de meilleures prévisions. L'évolution dans le QALY est plus lisse, la tendance à savoir que les meilleurs états reçoivent les QALY plus élevés est respectée et les erreurs sont stationnaires à leur moyenne (elles suivent la tendance d'un bruit blanc). À noter que ce modèle ne compte pas plus de 10 paramètres significatifs, une MAE de 13% et un pouvoir explicatif de 43%.

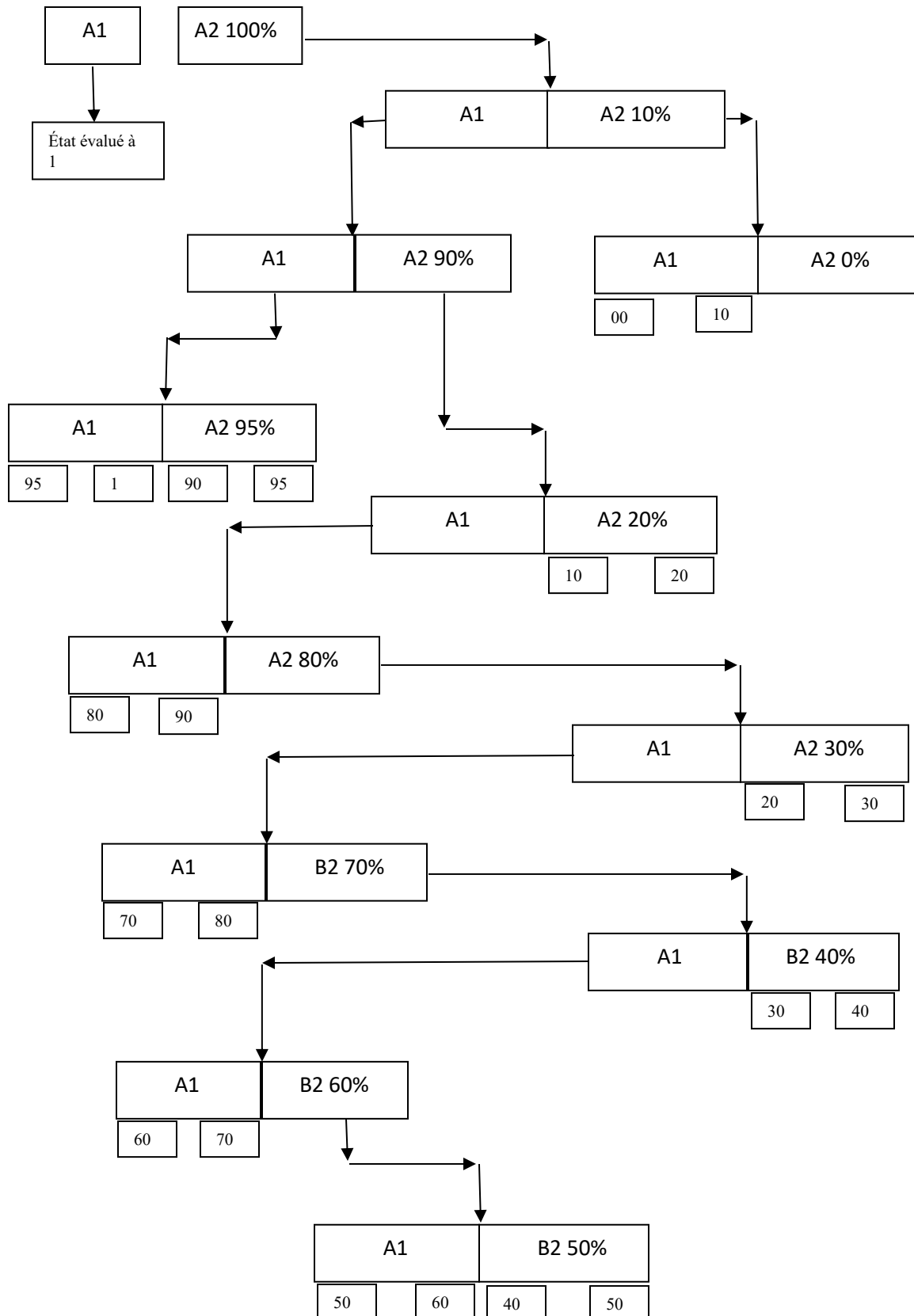
L'explication la plus simple qu'on peut apporter est que le QALY d'un état de santé dépend beaucoup plus des niveaux faibles des attributs que des meilleurs niveaux. En effet, quoi que l'on puisse dire, un traitement contient toujours un certain risque que ce soit le risque d'échec ou des effets secondaires. Donc que la valeur d'un traitement soit mieux corrélée à l'utilité des faibles niveaux permettant de mieux définir les états faibles qui justifie un traitement semble logique (on parvient à avoir au moins un pouvoir explicatif de 35% rien qu'avec la significativité des niveaux faibles). Le DCE parvient à trouver une significativité dans presque tous les niveaux. Le seul problème est de trouver la loi de probabilité adéquate afin de réduire les erreurs de mesures.

Finalement, le DCE associé au modèle de régression MCO par la moyenne est la meilleure alternative à la détermination des états du SF6Dv2 compte tenu des données de cette étude. Cependant, il n'est pas encore optimal et peut encore faire l'objet de grandes améliorations.

Les études à venir peuvent se concentrer sur l'hypothèse de la neutralité du risque dans les niveaux faibles (modifier le SF6dv2 ou la population à laquelle elle s'applique). Si la conclusion est favorable, on pourra par la suite se tourner sur une manière de jumeler le DCE avec la probabilité prise comme un attribut dans un modèle plus sophistiqué. Il sera par la suite question de trouver la loi de probabilité qui se rapprocherait au mieux du QALY tiré de ce DCE.

Une limite à cette étude est l'utilisation d'une enquête en ligne. Les réponses pouvaient être de meilleures qualités si l'enquête était conduite par des personnes expérimentées capables de mieux expliquer la tâche de chaque méthode et interagir avec les répondants en cas de besoins. La même limite a été soulignée par Norman et al (2014) concernant le SG.

## Annexe 1 : Arbre des valeurs associées aux états dans la méthode du SG



Cet arbre résume la procédure du pingpong et montre la manière dont est déterminé le QALY d'un état de santé. Ce QALY correspond au point où le répondant dit être indifférent entre un état intermédiaire qui lui est proposé et le traitement. Sur ce schéma ci-après, A1 représente un même état intermédiaire, celui que le répondant souhaite donner un QALY. On commence à lui proposer un traitement avec 100% de chance de succès. S'il refuse le traitement, on arrête le processus dans la mesure où son comportement signifie que A1 est au moins aussi bon que la santé parfaite qui est associée à un QALY égal à 1. Un tel comportement signifierait que la personne ne comprend pas le but de l'exercice ou il a une aversion pour le risque très élevé. Par contre s'il choisit le traitement le processus continue en réduisant les chances de succès à 10%. S'il le choisit, A1 a un QALY inférieur à 10%. Dans ce cas, on réduit davantage les chances de succès du traitement jusqu'à 0% de chance de succès. S'il refuse le traitement associé à 0% de chance de succès, on prend la valeur moyenne entre 0 et 10% comme étant le QALY de A1. S'il l'accepte, son comportement signifierait que A1 est pire que le PITS. Ce qui est impossible. On arrête le processus, on supprime cette réponse qui n'est pas cohérente, on remplace A1 par un autre état de santé intermédiaire dans la série de 5 à évaluer et le processus recommence. Par contre s'il avait refusé le traitement à 10% cela équivaldrait à dire que le QALY de A1 est compris entre 10% et 100%. Étant donné que l'intervalle est encore trop vaste, on fait varier les chances de succès du traitement en procédant par pingpong.

À partir de cet instant, on continue le processus si et seulement si :

- Le répondant refuse le traitement avec des chances de succès strictement inférieurs à 50% conformément au processus de Pingpong.
- Le répondant accepte le traitement lorsque celui-ci est proposé avec des chances de succès strictement supérieur 50% conformément au processus du Pingpong.

On arrête le processus si et seulement si ;

Le QALY est trouvé c'est-à-dire :

- Il choisit le traitement lorsque les chances de succès sont inférieures ou égales à 50%.

Par exemple s'il choisit le traitement avec 30% de chance de succès alors qu'il l'avait refusé avec 20%, A1 dispose d'un QALY situé entre 20 et 30%. En général, on prend la moitié c'est-à-dire 25%.

- Le répondant refuse le traitement lorsque les chances de succès sont supérieures ou égales à 50%.

Par exemple, il refuse le traitement avec une probabilité de succès à 70% alors qu'à 80% il le choisit, le QALY pour A1 sera dans ce cas compris entre 70 et 80% c'est à dire 75%.

## Annexe 2 : Description des niveaux des 6 dimensions du SF-6Dv2

SF-6D items	Descriptions
<i>Fonctionnement physique (PF)</i>	
PF1	<b>Pas du tout limité</b> dans les activités exigeant un <b>effort physique important</b>
PF2	<b>Un peu limité</b> dans les activités exigeant un <b>effort physique important</b>
PF3	<b>Un peu limité</b> dans les activités exigeant un <b>effort physique modéré</b>
PF4	<b>Très limité</b> dans les activités exigeant un <b>effort physique modéré</b>
PF5	<b>Très limité pour prendre un bain ou m'habiller</b>
<i>Limitations des activités (RL) : combinaison du rôle physique et moral</i>	
RL1	<b>Jamais accomplir moins de choses</b> que vous l'auriez voulu dans vos activités quotidiennes
RL2	<b>Rarement accomplir moins de choses</b> que vous l'auriez voulu dans vos activités quotidiennes
RL3	<b>Parfois accomplir moins de choses</b> que vous l'auriez voulu dans vos activités quotidiennes
RL4	<b>La plupart du temps accomplir moins de choses</b> que vous l'auriez voulu dans vos activités quotidiennes
RL5	<b>Tout le temps accomplir moins de choses</b> que vous l'auriez voulu dans vos activités quotidiennes
<i>Fonctionnement social (SF)</i>	
SF1	<b>Activités sociales jamais limitées</b>
SF2	<b>Activités sociales rarement limitées</b>
SF3	<b>Activités sociales parfois limitées</b>
SF4	<b>Activités sociales la plupart du temps limitées</b>
SF5	<b>Activités sociales tout le temps limitées</b>
<i>Douleur (PAIN)</i>	
PAIN1	<b>Aucune douleur</b>
PAIN2	<b>Douleurs très légères</b>
PAIN3	<b>Douleurs légères</b>
PAIN4	<b>Douleurs moyennes</b>
PAIN5	<b>Douleurs intenses</b>
PAIN6	<b>Douleurs très intenses</b>
<i>Santé mentale (MH)</i>	
MH1	<b>Jamais très nerveux ou triste et démoralisé</b>
MH2	<b>Rarement très nerveux ou triste et démoralisé</b>
MH3	<b>Parfois très nerveux ou triste et démoralisé</b>
MH4	<b>La plupart du temps très nerveux ou triste et démoralisé</b>
MH5	<b>Tout le temps très nerveux ou triste et démoralisé</b>
<i>Vitalité (VIT)</i>	
VIT1	<b>Jamais épuisé et vidé</b>
VIT2	<b>Rarement épuisé et vidé</b>
VIT3	<b>Parfois épuisé et vidé</b>
VIT4	<b>La plupart du temps épuisé et vidé</b>
VIT5	<b>Tout le temps épuisé et vidé</b>



**Annexe 3: Tableau de régression DCE (modèle agrégé à la moyenne)**

	MCO moyenne
PF2	-0,0196 [0,0621]
PF3	-0,0131 [0,0614]
PF4	-0,0853 [0,0615]
PF5	-0,137** [0,0637]
RL2	-0,112* [0,0611]
RL3	-0,0431 [0,0616]
RL4	-0,0898 [0,0610]
RL5	-0,133** [0,0624]
SF2	-0,0101 [0,0571]
SF3	-0,052 [0,0570]
SF4	-0,0218 [0,0567]
SF5	-0,0898 [0,0582]

PAIN2	-0,0619 [0,0605]
PAIN3	0,0198 [0,0611]
PAIN4	-0,085 [0,0614]
PAIN5	-0,166** [0,0689]
PAIN6	-0,225*** [0,0668]
MH2	-0,0368 [0,0549]
MH3	-0,117* [0,0595]
MH4	-0,157*** [0,0582]
MH5	-0,251*** [0,0579]
VIT2	-0,0277 [0,0562]
VIT3	-0,0657 [0,0574]
VIT4	-0,113** [0,0561]
VIT5	-0,151** [0,0579]

_cons	0,922***
	[0,0931]
N	120
PS	10
INC.	1
AIC	-30,464
BIC	42,01
MAE	0,14
JB	0,39
r2	0,478

Standard errors in brackets

\*  $p < 0,10$ , \*\*  $p < 0,05$ , \*\*\*  $p < 0,01$

Les nombres en bleu illustrent des incohérences.

### Bibliographies

John E. Ware, Jr., and Barbara Gandek. (1998). Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. *J ClinEpidemiol* Vol. 51, No. 11, pp. 903–912, 1998

Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. *J Health Econ.* 2011; 31:306 -18.

Brazier J, Usherwood T, Harper R, Thomas K. (1998). Deriving a Preference-Based Single Index from the UK SF-36 Health Survey. *J ClinEpidemiol* Vol. 51, No. 11, pp. 1115–1128

Brazier, J., Roberts, J. et Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*, 21, pp. 271-292.

Brazier J.E, Roberts J. (2004). The Estimation of a Preference-Based Measure of Health From the SF-12. *Medical care* 42(9):851-9

Craig B.M., Pickard S., Stolk E. et Brazier J.E. (2013). US Valuation of the SF-6D. *Medical Decision Making*. August: 793-803.

Ferreira L.N., Ferreira P.L., Pereira L.N., Rowen D., Brazier J.E. (2013). Exploring the Consistency of the SF-6D. *Value in Health*, 16, pp. 1023-1031.

Furlong WJ, Feeny DH, Torrance GW, et al. The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Ann Med* 2001; 33:375–84.

Furlong W, Feeny D, Torrance GW, et al. Guide to Design and Development of Health State Utility Instrumentation. Centre for Health Economics and Policy Analysis Paper 90–9. Hamilton, ON: McMaster University, 1990.

HAS. (2011). Choix méthodologiques pour l'évaluation économique à la HAS. Haute Autorité de Santé, 83 p.

- Karni E. *A theory of medical decision making under uncertainty.* *J Risk Uncertainty.* 2009; 39:1–16.
- Lancsar E, Louviere J. (2008). *Conducting discrete choice experiments to inform healthcare decision making: a user's guide.* *Pharmacoeconomics.* 26(8):661–77.
- Norman R, Viney R, Brazier J, Burgess L, Cronin P, King M, Ratcliffe J, Street D. (2013). *Valuing SF-6D Health States Using a Discrete Choice Experiment.* *Medical Decision Making,* 34(6):773-786
- Robinson A, Spencer A, Moffatt P. (2014). *A Framework for Estimating Health State Utility Values within a Discrete Choice Experiment: Modeling Risky Choices.* *Medical Decision Making,* pii: 0272989X14554715.
- Starmer C. *Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk.* *J Econ Lit.* 2000; 38:332–82.
- George W. TORRANCE. (1986). *MEASUREMENT OF HEALTH STATE UTILITIES FOR ECONOMIC APPRAISAL.* *Journal of Health Economics* 5 (1986) 1-30. North-Holland
- Torrance GW, Feeny DH. (1989). *Utilities and quality adjusted life years.* *Int J Technol Assess Health Care* 5:559–75.
- Van Houtven G, Reed Johnson F, Kilambi V, Brett Hauber A. (2014). *Eliciting Benefit–Risk Preferences and Probability-Weighted Utility Using Choice-Format Conjoint Analysis.* *Medical Decision Making,* 31(3):469-80.
- Samer A. Kharroubi, John E. Brazier, Jennifer Roberts, Anthony O'Haganb. (2006). *Journal of Health Economics* 26 (2007) 597–612.
- Amiram Gafni and Stephen Birch. (1993). *Economics health and health economics: HYE vs QALYs.* *Journal of Health Economics* 11 (1993) 325-339. North-Holland.
- Julie Ratcliffe, Leah Couzner, Terry Flynn, Michael Sawyer, Katherine Stevens, John Brazier and Leonie Burgess. (2011). *Valuing Child Health Utility 9D Health States with a Young Adolescent Sample.* *Appl Health Econ Health Policy* 2011; 9 (1): 15-27 1175-5652/11/0001-0015/\$49.95/0.
- Benjamin M. Craig, Jan J.V. Busschbach, Joshua A. Salomon. (2008). *keep it simple: Ranking health states yields values similar to cardinal measurement approaches.* *Journal of Clinical Epidemiology* 62 (2009) 296e305
- Donna Rowen, PhD, John Brazier, PhD, Ben Van Hout, PhD. (2014). *A Comparison of Methods for Converting DCE Values onto the Full Health-Dead QALY Scale.* *Medical Decision Making Web*
- ROSALIE VINEY, RICHARD NORMAN, JOHN BRAZIER, PAULA CRONIN, MADELEINE T. KING, JULIE RATCLIFFE and DEBORAH STREET. (2014). *AN*

*AUSTRALIAN DISCRETE CHOICE EXPERIMENT TO VALUE EQ-5D HEALTH STATES.* *Health Econ.* 23: 729–742 (2014).

Angela Robinson, MSc, Anne Spencer, PhD, Peter Moffatt, PhD. (2014). *A Framework for Estimating Health State Utility Values within a Discrete Choice Experiment: Modeling Risky Choices.* *The Medical Decision Making Web site*

Milton C. Weinstein, PhD, George Torrance, PhD, Alistair McGuire, PhD. (2009). *QALYs: The Basics. Volume 12 • Sup*

Han Bleichrodt, Jaco van Rijn and Magnus Johannesson. (1999). *Probability Weighting and Utility Curvature in QALY-Based Decision Making.* *Journal of Mathematical Psychology* 43, 238-260 (1999)

Han Bleichrodt. *A new explanation for the difference between time trade-off utilities and standard gamble utilities.* *Health economics* 11: 447-456 (2002)

Herbert E. Klarman, John O'S. Francis and Gerald D. Rosenthal. (1968). *Cost Effectiveness Analysis Applied to the Treatment of Chronic Renal Disease.* *Medical Care* 1968, Vol 4, No 1.

Richard Zeckhauser and Donald Shepard. (1976). *Where Now for Saving Lives?* Vol. 40, No. 4, *Valuing Lives* (autumn, 1976), pp. 5-45.

Menzies NA, Salomon JA. (2011). *Non-monotonicity in the episodic random utility model.* *Health Economics.* 2011; 20 (12):1523–31.

PETER WAKKER and ANNE STIGGELBOUT. (1995). *Explaining Distortions in Utility Elicitation through the Rank-dependent Model for Risky Choices.* (*Med Decis Making* 1995; 15:180-186).

Saroj Saigal, Janet Pinelli, Lorraine Hoult, M. Marie Kim and Michael Boyle. (2003). *Psychopathology and Social Competencies of Adolescents Who Were Extremely Low Birth Weight.* *PEDIATRICS* Vol. 111 No. 5 May 2003.

Zafar Hakim and Dev S. Pathak. (1999). *Modelling the EUROQOL data: A comparison of a discrete choice conjoint and conditional preference modelling.* *Health Econ.* 8: 103–116 (1999).

Amiram Gafni. (1994). *The Standard Gamble Method: What Is Being Measured and How It Is Interpreted.* *HSR: Health Services Research* 29:2 (June 1994).

Paul E. Green and V. Srinivasan. (1990). *Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice.* *Journal of Marketing* Vol. 54, No. 4 (Oct., 1990), pp. 3-19